Cost Modelling Using Automobile Warranty Data

by

Raymond Summit

A thesis submitted at Victoria University of Technology in fulfilment of the requirements for the degree of Doctor of Philosophy.

School of Computer Science and Mathematics

Victoria University

Melbourne, Australia

August 2004

Disclaimer

This thesis contians no material that has been accepted for the award of any other degree or diploma in any university or tertiary institution. To the best of my knowledge and belief, it contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

I confirm that this thesis does not exceed 100,000 words (excluding the bibliography and appendices).

Raymond Summit

August 2004

Abstract

This thesis sets out to model, from the manufacturer's point of view, the warranty cost of a repairable product. The product can be a complex one made up of numerous components, all of which are replaced upon failure and are non-repairable. The warranty-cost model is used to extrapolate the cost when the warranty is extended. Both point and interval estimates of the current and extended warranty costs are evaluated in this study.

The modelling in this thesis is based upon real data obtained from an Australian car manufacturer. As such, the thesis starts out with a detailed discussion of the important issue of data checking and cleaning.

Survival methods are used to model the product as a repairable system, with each repair consisting of the replacement of one or more failed components. Thus, the repair is taken to be as good as new, and the repair process can be modelled as a renewal process. Of interest is the expected number of replacements of each component during the warranty period, which can be estimated using renewal theory. To make the modelling manageable, the failure of each component is taken to be independent of the failure of other components.

The reliability of components with a small number of claims is assumed to follow the exponential distribution, whilst for components with more claims, the Weibull distribution is used. Although other models are considered, the exponential proves to be adequate when the number of claims is small compared to the number of items produced. Because of its versatility, the Weibull distribution is an appropriate choice when modelling components with a larger number of claims. Log likelihood methods are used to estimate the parameters of the models, from which the number of renewals during the warranty period are estimated. Numerical methods are employed to do this for the Weibull model.

The expected warranty cost for each component is calculated from the expected number of replacements and the expected cost of repair. The cost of repair is taken to be a variable quantity in this study. Using the variance of the expected number of renewals and the variance of the cost of repair, the variance of the warranty cost is obtained. The estimated warranty costs and variances of all components are used to obtain an expected warranty cost per vehicle produced, and a confidence interval on that cost.

The variance of the Weibull model proves to be too big to be of practical use. However, the exponential model's variance is quite useful. Simulation has been used to obtain a better confidence interval for the Weibull model. Simulations are also used to verify the results obtained by the modelling used in this thesis.

S-Plus has been used extensively throughout the thesis to perform the analysis. Although the number of functions in the S-Plus library is quite comprehensive, many new functions have been written by the author to undergo the analysis needed for this study.

Acknowledgements

I would like to gratefully acknowledge the assistance of my principal supervisor, Associate Professor Peter Cerone, Head of the School of Computer Science and Mathematics. Without his encouragement to apply for candidature and a scholarship, this project would not have begun; without his guidance and patience it would not have progressed to completion. I would also like to express my appreciation to Associate Professor Neil Barnett, who, as Head of the School, awarded me a School scholarship in the later part of my studies. I also appreciate the facilities provided at the School to enable me to conduct this project. Without these financial and other assistance, I would not have been able to undertake this project.

I gratefully appreciate the assistance of my co-supervisor, Dr Neil Diamond, whose assistance with statistical and S-Plus matters has been invaluable. Again, without his assistance, this project would not have come to completion.

I would also like to thank the Warranty Manager of the car manufacturer that has made the data available for this study. His generosity is much appreciated. Obviously without that data, this project would not have been possible. He and his company shall have to remain nameless so that the data is not misused against the company.

I would also like to thank my wife, Isobel, for her assistance in proof-reading, and also for providing moral and physical support throughout this project. I would like to dedicate this work to her and my children Katherine and Jessica, each of whom have had to go with only a limited amount of my time and attention during this project. I would also like to dedicate this work in loving memory of my parents, Mary and Charles, who encouraged me when a child, to learn and inquire.

Contents

Chapter 1
Thesis Introduction

1.1 Scope and Significance of This Study

Expensive, complex products are almost always sold with a warranty, so the cost of servicing a warranty is significant. Wasserman (1992) suggested that warranty claims could amount to 10-30% of production costs. Menezes and Quelch (1990) claimed that warranties represent an increasing cost to the manufacturer. They reported that the automotive industry in the United States of America spent over US$5 billion on product warranty in 1988, up from just over $700 million in 1965. Majeske and Herrin (1998) stated that in 1992, the combined total warranty payment bill for Ford, General Motors and Chrysler was $9.2 billion.

The cost of servicing warranty claims is to be paid for by the manufacturer, who must cover this cost in the sale price of the product. The most important reason for determining the cost of a warranty is the need to price the warranty (Hill and Blischke, 1987). Analysis of warranty data provides many benefits: it enables an estimate of the warranty cost to be made; it can predict the new costs if the terms of the warranty are altered; and it provides feedback to engineers about the reliability of the product, which may be valuable in reviewing the design of a product or the manufacturing process.

Manufacturing in Australia has become more competitive in the world market with the reduction or removal of import tariffs. The building of cost-effective, quality products has become an aim of many Australian (and international) companies. This is clearly exemplified in the car manufacturing industry, where the decrease of import duties has resulted in quality, imported vehicles becoming more competitively priced. Local manufacturers have had to build a reliable product at a cost-effective price to compete in this fierce market place. Over recent times, the standard warranty that comes with Australian-built cars has risen from one year to three years, but in contrast, many importers currently offer a more generous five-year warranty. Since the extent of a warranty is often used to judge the reliability of a product, it can be a powerful marketing tool.

The modelling of warranty costs is a complex study because of the variety of warranty terms, and because of the stochastic nature of product or component failures. Singpurwalla and Wilson (1993) affirmed that "the warranty problem is multidisciplinary, involving topics as diverse as economics, game theory, law, marketing, operations research, psychology, probability, and statistics."

One of the most difficult warranties to analyse is the two dimensional warranty, which is limited by both calendar time and usage. Although some literature has appeared since Singpurwalla and Wilson (1993) declared that "not much has been done with regard to two-dimensional warranties", there still appears to be a need for further work with two-dimensional warranty modelling. Although the data used in this thesis is two-dimensional, the approach used is a one-dimensional approach. This approach, as discussed in Subsection 5.2.8, is suitable for the terms of the manufacture's warranty. This simplified approach has been used in this thesis because of the large number of components to be analysed.

As much has already been written about the modelling of warranty costs, one might consider the subject to be in no need of further attention. However, most papers have taken the form of a theoretical treatise, with few studies investigating the application of the models. Blischke (1990) stated:

> *Perhaps more pressing, however, is the need for practical, applications-oriented research. Before they find widespread use in practice, the elegant models that have been developed must be incorporated into approaches that identify all significant cost factors and the associated data requirements, that emphasize the use of information that could realistically be attainable and that realistically model operational warranty progress. Methodological papers along these lines have not appeared in the statistical or management science literature (perhaps not because such studies have not been done, but because the results are proprietary).*

Thirteen years later, Jablonowski (2003) is still echoing these sentiments. He talks about the widening gap between theoretical developments and practice in risk management modelling, and what needs to be done to reduce the gap:

> *The existence of a wide gap, however, may be indicative that theoretical develop-*

ment in a field is not fruitful. ... The key is moving formulas and equations from paper to practice. ... Rewarding practical success, not just formal rigor, would go a long way to help assure that proper attention is paid to the development of "application friendly" theory. The corresponding threat is that without practical justification, or at least its promise, the theory will not be taken seriously.

This thesis attempts to reduce the gap between theory and practice. It uses real data from an automobile manufacturer's warranty database. Practical issues that concern the handling of real data are discussed in this study, from the methodology involved in the checking and cleaning of the data, to the implementation of theoretical models. These practical issues are an important aspect of this thesis, and therefore much effort has been devoted to them in this study.

Murthy and Blischke (2000) acknowledged that "For most companies, warranty costs are a closely guarded secret, as evidenced by the fact that very little warranty data is available in the public domain." In the quote of the previous paragraph, Blischke (1990) pointed out the proprietary nature of warranty data. Some studies have analysed car-warranty data, for example, Kalbfleisch, Lawless and Robinson (1991), Lu and Vance (1997), Majeske and Herrin (1995 and 1998), and Lu (1998). However, because of space limitations, these papers have discussed the analysis of one or two parts only. There is a need for a more extensive study that includes the warranty cost covering all components of a complex product, such as the automobile. There is a need to explore the processes involved in a larger study, and there is a need for practically-based research. In addition, the literature appears not to contain any warranty data from an Australian manufacturer. This thesis models the warranty cost of an entire vehicle manufactured in Australia. It also discusses in detail the practical issues that need to be resolved when analysing the real data.

It is common practice in the literature to assume that the cost of repairing a product is a constant. However, a manufacturer can only indicate a range of costs that it is willing to pay to the repair agent for particular repairs, rather than one particular price. The data obtained for this study reveals that the cost of replacing a particular component resulting from a warranty claim is far from constant. By allowing the cost of repair to be a variable, this study extends the current warranty models.

The current literature on the use of simulations in warranty-cost modelling is complemented by the use of simulation to verify the models that are established in the this study, and by the use of simulation to validate the techniques used in developing the models. Some of the simulations produce new results that complement the current literature on warranty simulation.

An integral part of this thesis is the extensive use of S-Plus in developing the many functions needed to complete the analysis. Although the S-Plus library of functions is vast, new functions have been developed to do the processing required in this study.

To summarise, this thesis extends the current warranty-cost models in the literature by presenting a practically-oriented, extensive study on warranty-cost modelling that uses an Australian automobile manufacturer's entire database for one year's production. The database contains warranty data for the entire length of the warranty of all of these vehicles. The aim of this thesis is to use this database to model the manufacturer's cost of a warranty, and then to use the model to predict the cost when the warranty terms are extended. Confidence intervals, often not discussed in the literature, have also been estimated in this dissertation. A further extension to the current literature is achieved by treating the cost of repair as a variable rather than as a constant, and by presenting simulations that verify the results of the modelling, and that validate the techniques used in the modelling.

1.2 Thesis Outline

This first chapter has identified the need for the current study, and has established its significance in the literature.

The next chapter discusses some perspectives on warranties. The purpose of that chapter is to provide some background to the modelling of warranty costs.

Chapter 3 explores the way in which the warranty process can be modelled. Included is a detailed discussion of the modelling of the two-dimensional warranty. The chapter also discusses the various approaches to warranty-cost modelling that has appeared in the literature.

The manufacturer's data is dealt with in Chapter 4. It discusses the practical issues of managing the real database, including how the data is imported into S-Plus, what exploratory data analysis is conducted, how errors in the data are detected and cleaned, and how the survival database is constructed.

The theory behind the models used in this dissertation is then considered in Chapter 5. It discusses the assumptions made, the survival analysis methods used, the non-parametric and parametric methods used to estimate the reliability of components, and the use of the renewal process in modelling component reliability. This is then put together to obtain the warranty cost of each component and then the overall warranty cost to the manufacturer.

Chapter 6 discusses how these theoretical models are used in conjunction with the cleaned data from Chapter 4, to obtain a specific model for the data. Again practical issues of joining the data to the theory are discussed in this chapter. Exponential and Weibull models are used to fit specific data from the manufacturer's database. Model estimates of the cost of the current warranty and an extended warranty are made, together with confidence intervals for these costs.

Chapter 7 consists of a number of simulations. A bootstrap simulation is presented to verify the results of the modelling in Chapter 6. Further simulations that generate data representing a whole year's production are presented: one for a three-year warranty, and one for a five-year warranty. A sensitivity analysis on the parameters of the model is performed, and then three more simulations are presented. These explore the fitting of a model to warranty data under various situations that mimic the collection of warranty data. The results of the simulations are used to validate the techniques used in modelling the data.

The thesis concludes with a discussion of the findings and limitations of this study, and identifies areas that could possibly be further explored.

A number of appendices have been included for completeness and ready reference. Contained in the appendices are a number of S-Plus scripts that have been used in the analysis of the data. These have all been written by the author of this thesis, specifically for this project. A feature of these functions is that they have been designed to handle input parameters, so that they can easily be reused with other data sets. It is also possible to adapt the functions to databases from different manufacturers.

Apart from the S-Plus scripts, a number of graphs are contained in the appendices. They have been included in this dissertation because they support the discussion in the main body, but have been placed in the appendices because there are too many to include in the body of the thesis.

A compact disc has been included with this thesis. It contains all the S-Plus scripts, functions and datasets that have been generated for this project. Some tables that have been created are too big to include in the appendices, so they have been referred to on the compact disc. All the files and functions are included so that it is possible to run the entire project from scratch. However, many of the data objects have taken a lot of processing time and have been assembled over several months. In order to be able to view the files, S-Plus 2000 or S-Plus 6 needs to be installed on the viewing computer. Two main folders can be found on the compact disc: Warranty and Warranty6. Shortcuts, which can be used to launch the project in S-Plus, called *Warranty* and *Warranty6* are included on the compact disc. This should copied onto the computer's desktop and used to start the S-Plus session, as it opens the project with the required objects for this project.

Chapter 2
Warranty Perspectives

2.1 Introduction

The previous chapter provided an outline of this thesis and identified the need for this study. The practical nature of this dissertation was pointed out. This study uses an automobile manufacturer's warranty database and discusses the detail of how this data is used to model the warranty cost of the manufacturer's warranty, and how this model is then used to estimate an extended warranty. Before starting the exploration and analysis of the data, a background to the study of warranty modelling is provided.

This chapter discusses a few different perspectives on warranty. The first perspective is a historical one. In the first section, the colourful history of warranties that has lead to the accepted present-day warranties is presented. A definition of a warranty is then given. A classification that is based on the terms of the particular policies follows this. Two-dimensional warranty policies, that are based on two attributes, such as time and usage, are then explored in their own right. Following this are two sections that explore the link that warranties have with management and engineering. Business decisions involving aspects such as planning, marketing, warranty servicing policies, warranty reserves, legal obligations, and warranty data management are discussed in the management section. The engineering section considers the effects on warranty cost of issues such as product design and development, quality control programs and product maintenance policies.

2.2 A History of Warranties

The beginning of this section is adapted from Loomba (1996).

Warranties can be traced back to the days of the Babylonians in the twenty-first century B.C. (Thomas, 1999). The eye-for-an-eye rule was one of the first policies imposed upon sellers and tradesmen. Examples of laws governing the trade of slaves and livestock are cited in Loomba (1996). The vendor had to provide a refund or replacement of goods if a known defect was hidden from the buyer, with a time limitation ranging from one day to one month. Examples in which the buyer had no redress are also cited.

Warranties based on moral and religious virtues can be found in ancient worlds, for example, in the Ancient Hindu (circa A.D. 500), Early Islamic (circa A.D. 632-661) and Jewish (circa second century A.D.) civilisations. They were also evident in Europe in the Middle Ages, where the powerful church denounced the selling of goods for profit. However, trade did exist and was accepted if it was not in the pursuit of profit, and the goods for trade were free of defects. Little, if any, comeback was available to the buyer of shoddy goods. In the early fourth century A.D., crafts became popular, but warranties were not needed as the merchants and products were locally known.

The general rule of *caveat emptor* ("let the buyer beware") applied during the industrial revolution. This worked reasonably well because products were simple and the consumer could understand and evaluate them before purchase (Blischke and Murthy, 1994). Vendors were locally known and word of mouth was often used to assess a product.

By the late nineteenth century in the U.S.A., standardised product warranties had emerged, but were very limited, due to the powerful position of the manufacturers. Deceit, misbranding, adulteration and misrepresentation were widespread. Often warranties were offered without any intention of fulfilling them and consumers perceived warranties to be an indication of poor quality. The growth of more complex products brought with it a greater need for consumer protection and the emergence of organisations that independently tested products. They were sponsored by insurance companies, underwriters and consumer-sponsored organisations. In 1914, the Federal Trade Commission in the U.S. established a set of codes for the selling of goods. Several versions of these codes were enacted by congress in the 1930s.

In the meantime, the U.S. courts began to make exceptions to the rule of *caveat emptor*. Kelley (1996) described a prime case. In 1939, Baxter challenged the Ford Motor Company in the courts, over its advertising claim that their vehicles had a triple shatterproof glass windshield. Baxter sustained injuries after a stone struck his windshield and shattered it. Ford argued that no warranty could exist without privity (that is, direct contractual relationship between buyer and seller), and that express warranties do not attach themselves to the product sold. The Washington Supreme Court disagreed with Ford.

By 1952, all states except Louisiana had adopted the Uniform Commercial Code. The code specified a manufacturer's obligations for both express and implied warranties. In express warranties, the emphasis was that the seller's promise to provide goods that fitted the promised description formed part of the deal. When an express warranty was not given, an implied warranty of merchantability applied to goods sold. The Commercial Code was primarily aimed at commercial transactions and was suppletory, not regulatory. The rationale for this was that merchants were expected to have sufficient knowledge and bargaining power to protect themselves in commercial transactions.

Consumers, on the other hand, did not have this expertise and bargaining power. Hence, there was a need for regulatory laws. The Magnuson-Moss Warranty Act was enacted in 1975 for this purpose. It followed ten years of studies, proposals and hearings aimed at overcoming warranty problems such as excessive use of disclaimers, inadequate coverage, consumer difficulty in obtaining warranty service and complex warranty language. The Act specified that a full warranty was to provide for the free replacement of defective goods for an unlimited duration, and was to include compensation for incidental damages. It also specified that a seller did not have to offer a full warranty, but could provide a limited warranty, where the terms and limitations were specified. Needless to say, many manufacturers changed their warranties to read "Limited Warranty" after the passing of this act. The act also specified that an implied warranty of merchantability was inferred in the sale of all goods, and could not be reduced by a limited warranty. The implied warranty stipulated that goods were to be "fit for ordinary purposes for which such goods are used" (Kelley, 1996).

According to Beerworth (1991), the most important sources of product liability rules in Australia are the tort of negligence (the common law) and Division 2A, Part V, Trade Practices Act 1974. The law of negligence imposes liability upon a person who does not use reasonable care in the manufacture and design of a product. Division 2A, on the other hand, imposes liability upon a manufacturer for a product which is unmerchantable or unfit for its stated purpose.

By the second half of the twentieth century, dramatic changes were taking place in the role and importance of warranties, in relation to product sales and service. Blischke and Murthy (1992) identified four main factors responsible for this. Firstly, consumers were becoming more aware of their rights as a result of consumer groups being more active and vocal. Secondly, governments were responding to the concerns of these groups and were legislating to protect their rights. Thirdly, manufacturers were reacting to pressures from both consumer groups and governments. Lastly, manufacturers were realising the importance of warranties as a marketing tool. (This is discussed further in Section 2.6.)

By the late 1980s, consumers were becoming more quality sensitive. Manufacturers responded to this demand by producing high quality products and backing them up with longer warranties. Menezes and Quelch (1990) identified three factors for this consumer focus on quality. These comprised the availability of quality products originating from Japanese companies; the fact that many consumers have high disposable incomes but little time or inclination to deal with product failure; and increased product complexity, often leaving the consumer unable to judge quality before buying the product.

By the 1990s, manufacturers have focussed on customer satisfaction. Warranties were seen as a means of assuring this satisfaction. For example, General Electric in the U.S. had a "satisfaction guaranteed" program for all its major appliances that allowed customers to return any appliance within ninety days, no questions asked (Menezes and Quelch 1990). Similarly, in Australia, O.P.S.M. currently offers a spectacles warranty which includes free exchange with glasses of comparable value during the first four weeks if customers are dissatisfied for any reason.

2.3 Definition of a Warranty

In simple terms, "a warranty is a producer's guarantee that a product or service will adequately meet performance requirements for a certain period" (Thomas, 1999). More specifically, it is that a warranty is a contractual agreement between a manufacturer, or seller, and a consumer, that a product is, or certain of its performance characteristics are, free from defects in materials and workmanship; it is a commitment to correct problems if the product fails during the warranty period (Menezes and Quelch, 1990, and Blischke and Murthy 1992 and 1994). If the product is not used under the specified normal conditions, the warranty is nullified, and the manufacturer's obligations are released. A manufacturer can also use a warranty to identify the limitations of liability when the product is used in the specified manner. For the consumer, a warranty is an insurance policy against product defects.

2.4 Classification of Warranties

The following classification of warranty policies is an adaptation of the taxonomy presented by Blischke and Murthy (1992 and 1994). The policies are not necessarily mutually exclusive.

1. *Free Replacement Warranty*: With this warranty, the manufacturer either replaces, repairs or reimburses the customer for a failed product for a period of time t_w commencing from the time of the initial purchase. This policy is the most common warranty on consumer goods, ranging from nonrepairable inexpensive products, such as film, to expensive repairable items, such as automobiles and refrigerators.

2. *Pro-rata Warranty*: In this warranty, a manufacturer agrees to refund an amount in proportion to the remaining time left on the warranty bounded by time t_w from the time of the initial purchase. This policy is appropriate for products such as automotive batteries and tyres, that wear out and must be replaced when they fail.

3. *Combined Free Replacement and Pro-rata Warranty*: Up to time t_{w1}, this policy is a free-replacement one. Between t_{w1} and t_{w2} (where $t_{w1} < t_{w2}$) the warranty converts to a pro-rata policy.

4. *Renewing Warranty*: Upon failure of a product during the warranty period, a replacement is made, and the warranty starts anew. This policy is offered with inexpensive electrical, electronic or mechanical products where the warranty is contained inside the product's packaging. By returning the new warranty registration card, the warranty starts anew.

5. *Fleet Warranty*: A group of N items are covered as a lot, for a total period of Nt_w. This type of warranty is applicable to components of industrial and commercial equipment bought in lots, as spares, and used one at time until failure. This policy was referred to as the fleet warranty by Berke and Zaino (1991), who pointed out that

the fleet warranty has advantages to the manufacturer over the combination policy because the effect of infant mortality is averaged out with longer wearing parts.

6. *Reliability Improvement Warranty*: This policy has evolved from airline companies purchasing commercial aircraft and is now mandatory for any U.S. Department of Defense purchase of military equipment (Nenoff, 1988). Under this policy, the manufacturer undertakes to repair or replace the product or its components within a specified turnaround time, and make design and engineering changes necessary to meet the required reliability, as measured by the mean time between failure. This type of warranty has also been referred to as the essential performance requirements warranty (Gilbertson, 1989).

The above policies can be restricted on one variable, such as calendar time or usage (measured, for example, by distance travelled or flying hours), or on two variables, such as time and usage. The two-dimensional warranty is usually offered in the automobile and aviation industries.

In any of these policies, replacement or repair upon failure is at the manufacturer or repairer's discretion, but may be written into the warranty policy. Replacement is appropriate for non-durable and inexpensive consumer goods, or where the cost of repair would be greater than the cost of replacement. Repair is more appropriate for expensive, durable goods.

When a manufacturer wants to limit the warranty by usage, but does not want to impose a two-dimensional warranty, he may offer different policies to different segments of consumers. For example, home users may be offered a two-year warranty on a washing machine, whilst an industrial or commercial user may be offered six months.

In the next section, two-dimensional warranty regions are categorised.

2.5 Two-Dimensional Warranty Regions

Some products or components age with both use, due to wear, and time, due to rusting, as well as being affected by humidity and other environmental factors (Singpurwalla and Wilson, 1993). In such circumstances, it is appropriate to limit the warranty by time and usage.

The two-dimensional warranty has been referred to under different names by various authors. Moskowitz and Chun (1994) and Chun and Tang (1999) called it the "two-attribute warranty". Farewell and Cox (1979) referred to multiple time scales, while Lawless (1995) and (1998) described it in terms of covariates. The term "two-dimensional warranty" is used in this dissertation.

Two-dimensional warranties can be subdivided according to the two-dimensional regions representing their policies. The following two-dimensional warranty policies are an adaptation of those presented in Blischke and Murthy (1992) and (1994), Murthy, Iskandar and Wilson (1995), Singpurwalla and Wilson (1993), and Chun and Tang (1999).

1. *Rectangular Region*–Figure 2.1(a): Coverage is for a maximum time t_w and usage u_w, whichever comes first. Cars and aeroplanes typically come with this type of warranty.

2. *L-shaped Region*–Figure 2.1(b): This policy is a compromise between Policy 1 and Policy 2 above. Under this policy, the product is covered for a minimum time period t_{w1} and for a minimum usage u_{w1}. The manufacturer's obligation is limited by a time t_{w2} and usage u_{w2}. Because of the time and usage restrictions, neither heavy users nor light users subsidise the other. Again, no product is known to be covered by this policy.

3. *Infinite L-shaped Region*–Figure 2.1(c): Coverage is for a minimum time t_w and minimum usage u_w, whichever occurs later. No product is known to be offered with this warranty. As failure of most products depends upon both usage and age to some extent, a manufacturer would be reluctant to offer this open-ended warranty.

4. *Triangular Region*–Figure 2.1(d): This policy is an extension of Policy 3, where a linear decrease from a maximum usage is used instead of a cut-off at u_{w2}. This policy covers a product up to a maximum time, t_w, from purchase and for a maximum usage of u_w. Conceptually, this is the fairest policy since neither heavy nor light users subsidise each other, and this is accomplished in a smooth way, avoiding the jumps in Policy 2. Again, no product is known to come with this warranty policy, perhaps because of the complexity for the lay-person to determine whether they are still covered by warranty. This could lead to customer dissatisfaction.

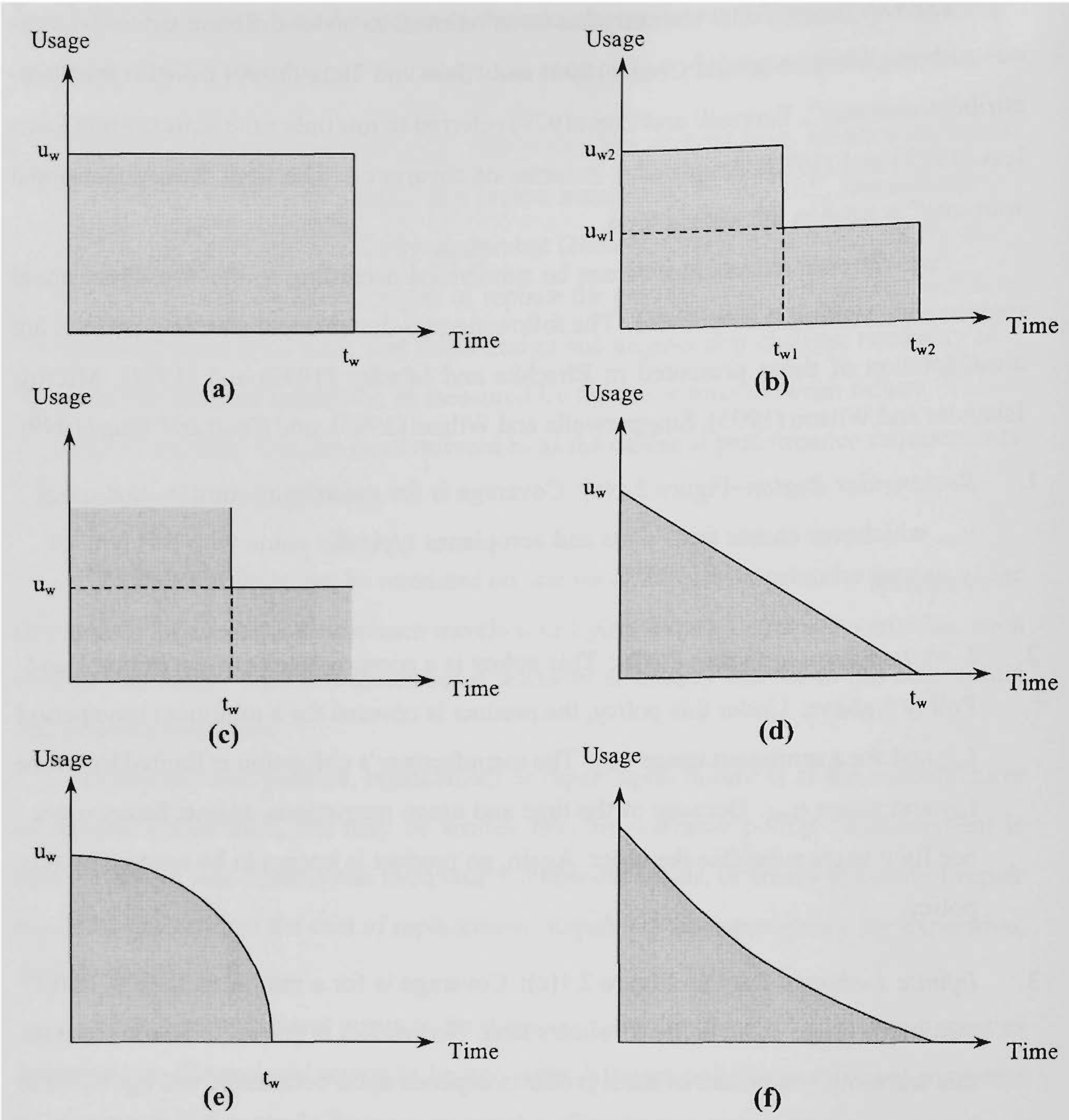


Fig. 2.1. *Two-dimensional warranty policies.*

5. *Circular Region*–Figure 2.1(e): This is another variation of the triangular region in an attempt to make the warranty policy fairer for the average user.

6. *Flexible Region*–Figure 2.1(f): This policy is referred to as the iso-cost policy by Chun and Tang (1999), and the flexible two-attribute policy by Moskowitz and Chun (1994). With this policy, a customer can choose, at purchase, any rectangular warranty plan represented as a point on the iso-cost curve. The iso-cost curve is

determined so that any point on the curve has the same expected warranty cost to the producer. Moskowitz and Chun (1996) and Gertsbakh and Kordonsky (1998) highlighted the appeal of this policy, as the flexibility may be attractive to a variety of customers.

Other variations of the two-dimensional region have been reported in the literature, but the above are probably the most practical. One variation is the unlimited distance warranty that is offered with the sale of some vehicles. This warranty is dependent on time only, and so is, of course, a one-dimensional warranty.

The rectangular region of Figure 2.1(a) is the one in most common usage in the automobile and aviation industry. It is simple to implement, and leads to little, if any, discrepancy.

For most other products, a one-dimensional warranty is offered, and consumers usually have a limited range of usage. Heavier users, such as commercial users, are generally offered a shorter time under warranty. Interestingly, in more recent years, this policy has been offered with imported cars in Australia. It favours the heavy user.

Many of the other policies of Figure 2.1 are attractive since they are fairer to all users. However, there do not appear to be any products offered with these other policies because they are more difficult to implement, more costly to administer, and more likely to lead to disputes and ill-feeling amongst customers.

A pro-rata variation on these two-dimensional policies is also possible. For example, with the Rectangular policy, the first part of this policy is identical to Policy 1, where the product is covered for a maximum time of t_{w1} or usage u_{w1}. The pro-rata period extends this to a maximum time t_{w2} or usage u_{w2} on a pro-rata basis. This is illustrated in Figure 2.2. Again, there do not seem to be any products covered with this warranty policy.

This study focuses on the two-dimensional, free-replacement warranty.

2.6 Warranty and Management

Blischke and Murthy (1992) identified the type of questions that a manufacturer needs to answer about warranties: What is the cost of a specific warranty? How do the competitors' warranties compare? How does the cost of warranty change with a change of parameters (for example, duration, form of rebate)? How does one optimise the choice of warranty

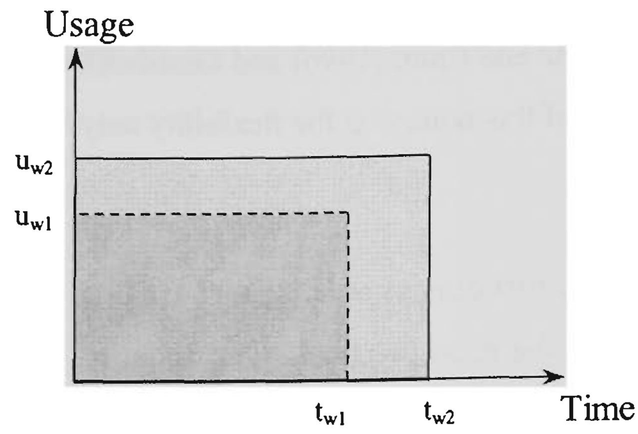


Fig. 2.2. *Two-dimensional pro-rata policy.*

when multiple business objectives are involved? What is the optimal strategy for servicing a warranty? What kinds of data are needed and how should the data be analysed? What are the optimal decisions with regard to product design and manufacture, given that the product must be sold with a specific type of warranty policy?

A number of publications have endeavoured to integrate the many aspects of warranty analysis. Most notable of these are Blischke and Murthy (1992) and (1994), Murthy and Blischke (1992a) and (1992b), and reviews by Hill and Blischke (1987), Blischke (1990) and more recently, Murthy and Djanaludin (2002).

The discussion in this section reviews some of the literature that explore various issues of warranty analysis.

2.6.1 Decision-Making

Management is faced with setting many production variables in the manufacturing process. Some studies have attempted to provide a management model based on various decision variables. For example, Mitra and Pantankar (1993) provided a model that integrates the marketing, production and financial variables into a multi-objective function. Decision variables included the price, warranty length, production quantity, and lot size. System constraints on the above decision variables were based on minimum and maximum values between which those variables should lie. Several goals were considered at different levels of priorities. Thus, optimisation of specific goals, such as cost minimisation or profit maximisation was made possible by adjusting the decision variables.

The life cycle approach has been used by some authors. For example, Murthy and Blischke (2000) discussed engineering (technical) issues and management (commercial) issues that affect a product throughout its life. A product's life-cycle can be divided into three broad phases. The prelaunch stage, which includes product design, development of a marketing plan and the specification of a warranty. The launch phase focuses on the target launch date. The postlaunch phase includes servicing the product, feedback and design revisions, and fulfilling the warranty obligations. By integrating warranty policy with other aspects of decision-making, the cost of a warranty can be well planned for.

Another example of a study using the product life cycle approach was the one by Polatoglu and Sahin (1998). It investigated the probability distributions of the manufacturer's rebate, cost, revenue and profit during a cycle under a renewing, combination free replacement/pro-rata warranty policy.

A manufacturer needs to estimate the warranty cost of a product so that it can be built into the purchase price. Blischke and Scheuer (1975) modelled the long-term profit to the producer when a product is sold with and without a warranty. Both the free-replacement and pro-rata warranties were discussed. In the same study, the long-term cost to a buyer was also modelled.

2.6.2 Marketing

A favourable warranty can reduce consumer-perceived risk and can be used as a marketing tool. For example, Chrysler in the United States extended its power train warranty from two years/24,000 miles to seven years/70,000 miles, and Chrysler aggressively advertised this fact. Industry analysts calculated that this resulted in Chrysler gaining at least one market share percentage point (Menezes and Quelch, 1990). Another example was Hyundai's extension in 1990, of its one-year warranty to three-years/60,000 kilometres. This attempt to gain a greater Australian-market share on its relatively new Excel resulted in a sales increase of 40% in the following year (Hyundai, 2004.)

As Murthy and Blischke (1992a) pointed out, manufacturers are sometimes forced to offer better warranties in order to keep up with the competition. A good example of this can be found in the automobile industry, where the trend to offer less favourable warranties in the sixties and early seventies was reversed after the penetration of the market by Japanese cars, which were offered with more favourable warranties. By the turn of the century, Australian manufacturers were offering 3 years/100,000 kilometres as a standard warranty because of pressure from importers.

In a highly competitive marketplace, manufacturers see customer satisfaction as important. Murthy and Blischke (2000) pointed out that product support, through the servicing of a warranty, is important in fostering customer good-will, and can lead to subsequent sales. According to Stahl and Grigsby (1997, page 165), Ford estimates that it costs five times as much to attract a new customer as it does to retain an old one.

Menezes and Quelch (1990) argued that from a manufacturer's perspective, customers that experience a problem with their product should be encouraged to invoke the warranty for two reasons. One is that consumers with problems often do not complain, they just switch brands. A second reason is that a dissatisfied customer has the potential to do harm by spreading strong negative word-of-mouth comments about the product. A customer that feels satisfied with problem resolution is more likely to be loyal to that brand and spread positive word-of-mouth communication.

Elsayed (1996, page 475) summed it up as follows:

> *The increasing worldwide competition is prompting manufacturers to introduce innovative approaches in order to increase their market shares. In addition to improving quality and reducing prices, they also provide attractive warranties for their products. In other words, warranties are becoming an important factor in the consumer's decision-making process. For example, when several products that perform the same function are available in the market and their prices are essentially equal, the customer's deciding factor of preference for one product over the other includes the manufacturer's reputation and the type and length of the warranty provided with the product. Because of the impact of the warranty on future sales, manufacturers who traditionally did not provide warranties for some products and services are now providing or required to provide some type of warranty.*

2.6.3 Warranty Servicing Policies

Some studies have analysed the costs of different warranty servicing policies. For example, Dagpunar (1992) compared the cost of two warranty servicing policies. The first policy is an adaptive repair-cost limit, which is set dynamically according to the age of the product and the length of warranty remaining. The second is a constant repair-cost limit throughout the warranty period. Dagpunar concluded that there was little cost benefit in the adaptive policy, especially when the additional administrative costs of the policy are taken into consideration.

In the case of automobile warranty claims, repair is almost certainly the only option, as replacement of the entire vehicle would be unnecessary and far too expensive. Thus, the above strategies are not used in this thesis.

2.6.4 Warranty Reserves

A number of studies have modelled the size of the warranty reserve needed to honour a warranty. For example, Menke (1969) considered a nonrepairable product sold with a pro-rata warranty that had an exponential failure distribution. Amato and Anderson (1976) factored into the model the present-day value of warranty claims and changes in the price level. Thomas (1989) extended the warranty reserve models by considering the uniform, gamma and Weibull failure distributions. Eliashberg, Singpurwalla and Wilson (1997) made further extensions by modelling warranty reserves for two-dimensional warranties. Jun and Pham (2004) modelled the warranty reserve for the free-replacement and pro-rata warranty policies when the value of money in discounted over time and the repair in minimal on a series system.

In the current study, the warranty cost per vehicle is estimated, from which a warranty reserve can then be calculated.

2.6.5 Legal Obligation

Legal liability is a cost that the manufacturer may have to meet. Morgan (1982) stated that the cost of liability from product warranties has been increasing over the years.

Blischke and Murthy (1994) identified the following areas for which a manufacturer can be held liable: negligence, including breach of duty of reasonable care with regard to design, manufacture, inspection and the warning of any foreseeable danger; the quality of the product, including any physical harm or injury caused by a product used according to the manufacturer's instructions; misrepresentation, if unrealistic performance claims are made about a product, be they false representation, or based on laboratory or consumer research; and breach of warranty, when a manufacturer does not honour the terms of the warranty.

2.6.6 Data Collection and Management

The importance of data collection and management was pointed out by Lawless (1998, page 41):

> *If constructed and maintained properly, warranty data bases may be used to predict future claims, to compare claims experience for different groups of products, and to study variations in claims relative to factors such as time and place of manufacture, or usage environment. In some circumstances warranty data may also be used to estimate the field reliability of products and to identify opportunities for the improvement of quality and reliability.*

2.7 Warranty and Engineering

A manufacturer's warranty cost depends upon the reliability of the product. The reliability of a product, in turn, depends largely upon its design characteristics and the ability of the manufacturing process to produce an item according to the specified design. Greater reliability is achieved by improvements in the design and development of a product, and improvements in quality control during the manufacturing process. These improvements, however, come at a cost, and an engineering decision has to be made as to where the balance between these competing expenses should lie. Studies that explore the connection between warranty cost and engineering design and manufacture are discussed in this section.

2.7.1 Manufacturing Stages

Blischke and Murthy (1994, Chapter 10) presented a comprehensive discussion of the relationship of warranty cost and various engineering decisions. The effects on warranty cost of engineering decisions made at the design stage, the manufacturing stage and pre-sale testing stage were explored. Blischke and Murthy developed a manufacturing cost model that was dependent on warranty and manufacturing costs, both of which, in turn, were dependent on the parameters of the reliability of the product. These parameters were considered to be a variable that can be adjusted by the engineer, and which can take on values within a given range.

2.7.2 Quality Control

The study of warranty data provides an opportunity to monitor the reliability of a product. As Suzuki (1985) put it, a warranty "serves not only the owners, but also the manufacturers who wish to monitor the reliability of their products". Put in another way, "the ultimate test of a manufactured product is how well it performs in the field, that is, in the hands of the customers" (Lawless and Kalbfleisch, 1992). An example of a company that uses warranty data to monitor its manufacturing is given by Firestone, who use their monthly warranty analysis as an early warning system (Menezes and Quelch 1990).

Quality control is part of the manufacturing process and is achieved through inspection and pre-sale testing. Reducing batch sizes and increasing inspection increases the reliability of the end product, thus reducing warranty costs, but this is paid for in increased production costs. Many authors have discussed the issue of optimal batch size that minimises both production and warranty cost. For example, Chen, Yao and Zheng (1998) developed an end-of-production inspection procedure for batch-produced items. Their aim was to identify a checking policy that minimised warranty and inspection cost. The approach was to choose a sample from a batch of items. The batch was considered satisfactory if a threshold of defective items for a given sample size was not reached, and inspection was stopped. If the threshold was passed, further inspection was conducted, either until the threshold for the increasing sample size was met or the entire batch was inspected. In this way, a balance between inspection and repair cost and warranty cost was achieved. Wang and Sheu (2003) also discuss the trade-offs between manufacturing cost and warranty cost. They developed a cost model to find the optimal production lot size to minimise the total production and warranty cost.

Burn-in tests are used to find infant failures, improve the reliability of a product, and reduce warranty costs and customer dissatisfaction. This is particularly applicable to electronic components where infant mortality can be high. Kar and Nachlas (1997) used a renewal-theory approach to develop a net profit function that included burn-in cost, an in-plant component replacement cost, and a warranty cost. This non-convex function is optimised by using a sequential search over the extreme points of the feasibility region. The methods used in this study apply to decreasing hazard systems only, which limits the shape parameter of the Weibull model to less than one.

Warranty data can be used to monitor the reliability of components in a system. Lu's (1998) study was concerned with using early vehicle warranty data (4 or 5 months in service) to predict potential reliability problems. Such problems could have resulted in a product recall or other corrective action. Using data from Chrysler on a range of 23,365 cars of various models and 8,752 trucks, produced between 1983 and 1986, Lu modelled mileage accumulation rates on the lognormal distribution. The Weibull distribution was used to model the time to failure of vehicles. The reliability for each 1,000-mile intervals was calculated, and the data extrapolated to predict future claims. Early detection of reliability problems was also the subject of Wu and Meeker's (2002) study. They developed a procedure to detect potential problems from early warranty data, using both statistical decision rules and graphical methods. These techniques discussed in both of these studies would be very useful "quality control" information that could be used in the manufacturing process.

2.7.3 Predictive Model and Warranty Data

Majeske and Herrin (1998) compared the warranty cost predicted by a manufacturer's model against field data. They used two sets of warranty data to make their comparisons. The first set involved the redesigning of the radio to allow for new features, such as a compact disc player. The manufacturer used an exponential model and found that the warranty data showed a failure rate of twice that predicted by the manufacturer's model. Majeske and Herrin concluded that the Weibull model provided a better fit to the data than did the exponential model.

The second set of warranty data was based on the brake subsystem of a luxury car. The manufacturer redesigned some components within the system (for example, brake pads, rotors) to utilize state of the art materials. A Weibull model was used to predict the lifetime of the brake subsystem and it provided a good fit to the data. However, the model used by the manufacturer under predicted the number of warranty claims by a factor of four. Majeske and Herrin found that the bench test focused on inducing catastrophic failures that indicated the end of a brake component of subsystem life, and excluded other less severe failures, such as uneven wear in the brake pads and rotors. They concluded that manufacturers need to verify the predictive models obtained from bench testing against warranty data, and that the failure mode needs to be considered.

2.7.4 Maintenance Scheduling

The scheduling of preventive maintenance to minimise warranty costs has received attention by some authors. Chun (1992) included a scheduled preventive maintenance cost and a repair cost at failure for a product. His objective was to determine the optimal number of preventive maintenance operations during the warranty period in order to minimise the manufacturer's overall cost. He considered the case where preventive maintenance was imperfect, and repair at failure was minimal.

In a similar but more in-depth study, Lin, Zuo, Yam and Meng (2000) developed a cost model that incorporated warranty, periodic preventive maintenance and repair upon failure of a system. The study's aim, however, was to find the optimal design of a mixed series-parallel system. The overall cost, including manufacturing and set-up expenses, maintenance expenses and warranty costs was optimised.

In yet another study, Chen and Popova (2002) developed a maintenance policy to minimise servicing cost for an item with a two-dimensional warranty. An iterative procedure is used to estimate the item's failure rate distribution from warranty data, The authors used an algorithm based on Monte Carlo simulation to obtain the optimal maintenance policy.

In all three studies, the preventive maintenance cost was borne by the manufacturer. This may be the case with large military or civil projects, or with the reliability improvement warranty policy, but is generally not the case with consumer goods. Although there have been examples of car importers in Australia taking up the scheduled maintenance costs, it is generally not the practice, with the maintenance usually being paid for by the consumer. In this thesis, it is assumed that maintenance cost will be met by the consumer.

2.8 Conclusion

The purpose of this chapter has been to present some perspectives on warranties and thus provide some useful background to the study of warranty-cost analysis. Literature that discusses a range of warranty issues has been presented. In this chapter, the scope and significance of this study have been discussed. Various warranty policies have been classified according to the terms of the rebate. Two-dimensional warranties, which are of interest in this study, have also been classified. The links between warranty and management, and warranty and engineering have been explored.

The next chapter explores warranty modelling. It discusses the modelling of the warranty process, including two-dimensional warranties. The chapter also surveys current literature on the various approaches to modelling warranty costs.

Chapter 3
Warranty Modelling–A Review

The last chapter presented a broad overview of warranties. The purpose of this chapter is to enlighten on the breadth of approaches to warranty modelling that exist in the literature. The chapter starts with a consideration of the processes involved in modelling the warranty cost, from when a fault occurs, through to the cost of rectifying the problem. This is followed by a discussion of the modelling of two-dimensional warranties. The chapter ends with an exposé of a number of approaches used for product reliability and warranty costing.

3.1 Modelling the Warranty Process

The warranty process consists of item failure as perceived by the customer, the customer's response to that failure, and the manufacturer's or repairer's response to a customer's claim, if one is made. The following comprises a discussion of the modelling involved at each of these stages.

3.1.1 Modelling Failure Mode

Murthy and Blischke (1992b) identified two ways of modelling the time to failure: physically based or black-box based. The former approach is based on the physical nature of the failure, where items are considered to receive shocks of random magnitude at random time intervals. In the Shock Damage model, item failure occurs at the first instant that a shock exceeds a critical value. In the Cumulative Damage model, item failure occurs when the cumulative damage exceeds some critical value. In both cases, the time to failure is a random variable, and obtaining its distribution function involves the analysis of the stochastic process characterising the shocks.

In the black-box approach, the physics of the failure is not considered at all. The time to failure, X, is taken to be a random variable with a distribution function, $F(x)$, based on the modeller's physical judgement and on historical data.

Many complex products have a modular design. Each module is made up of a number of components, the failure of any of which results in module failure. Thus, to model failure, component failure needs to be related to module failure, and module failure needs to be related to item failure. Thus, the modelling of failures in a complex system can become very difficult. One simplification is to assume that the failure of a module does not affect the failure of other modules.

3.1.2 Modelling Time to First Failure

Item failures occur randomly over the time continuum. Thus item failures, in the case of one-dimensional warranties, can be treated as random points along a time axis. The two-dimensional case is discussed in Section 3.2.

Ascher (1992) pointed out that it is important to differentiate between the modelling of repairable and nonrepairable systems. It is appropriate to model the time to failure of a nonrepairable item or part with a suitable probability distribution function. Ascher suggested the approach outlined in Figure 3.1 (which is adapted from Figure 3 of Ascher) for analysing the inter-arrival times of a repairable system. For a repairable system, a distinction needs to be made between the first failure and subsequent failures of an item because any subsequent failure depends on the repair or replacement policy, as discussed in Subsection 3.1.5. In fact, early models of warranty costs, such as those of Menke (1969) and Amato and Anderson (1976), ignored the repair process altogether and were based on the time to first failure only (Hill and Blischke, 1987 and Blischke, 1990). Considering first failure only, the expected warranty cost of an item with a warranty length of time length t_w is given by

$$C(t_w) = c_s F(t_w),$$

where c_s is the cost to supply the item and $F(\cdot)$ is the probability distribution for the failure of the item.

The failure of components within a system is being modelled in this study. As such, their failure is being modelled by a suitable distribution function.

3.1.3 Modelling Customer Claim Behaviour

Not all failures lead to claims, as the buyer may decide to put up with a minor defect. If the product is subjected to regular maintenance, the failure may not be reported until the next scheduled service. Rai and Singh (2004) addressed the issue of the existence of spikes in warranty claims towards the end of a warranty period, and provided a methodology for

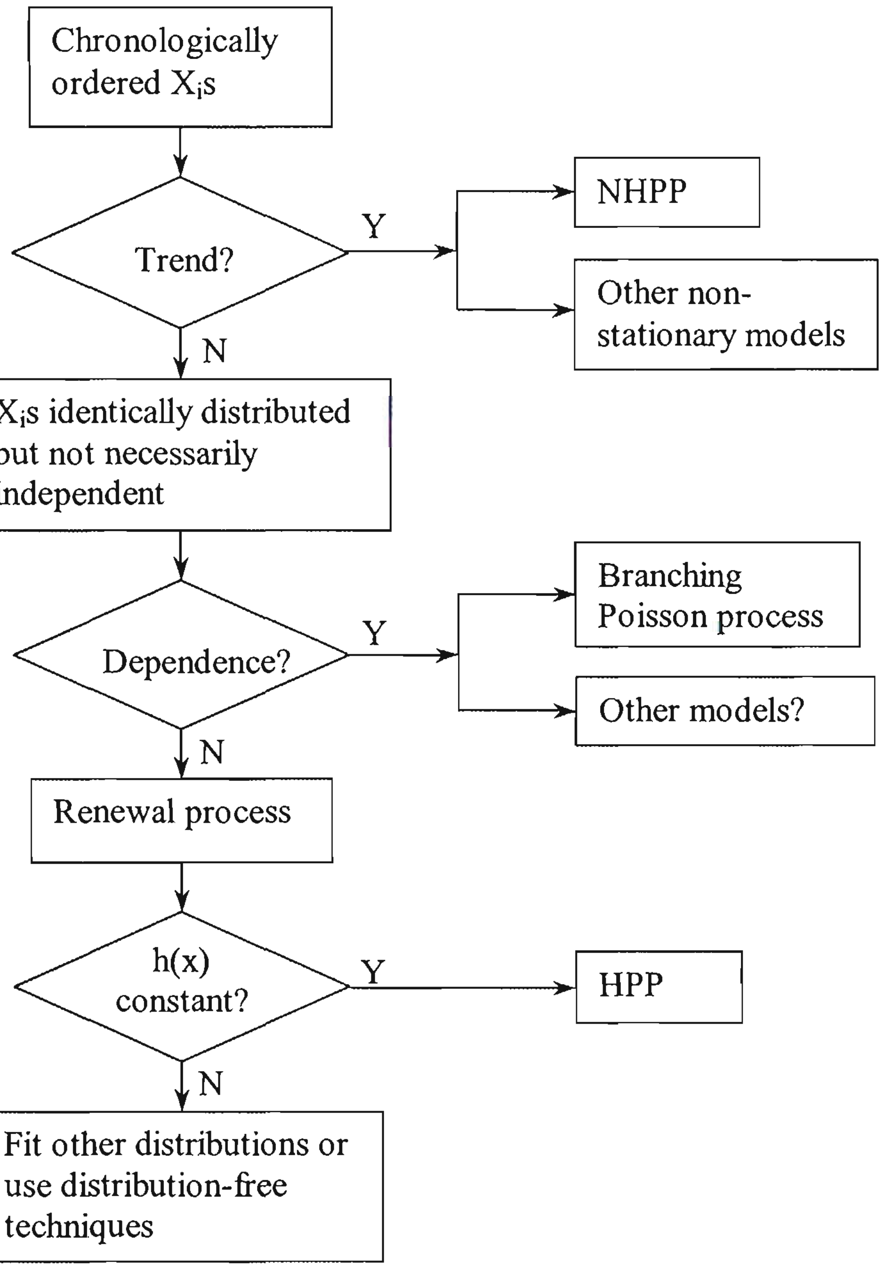


Fig. 3.1. *Statistical anlaysis of successive inter-arrival times of a repairable system.*

plotting the hazard function of components. This issue clearly puts another complexity on the modelling of failure times, as the reported times may not represent the true time of failure. Figure 3.2 shows that the spikes appearing in timing of claims for the data used in this study, although existent, are not very severe. Thus, their impact will be minimal. Indeed, from the manufacturer's point of view, modelling the claim time instead of the actual failure time is sufficient in terms of estimating the cost of the warranty. However, if the reliability of a component needs to be analysed, then the existence of claim spikes needs to be considered.

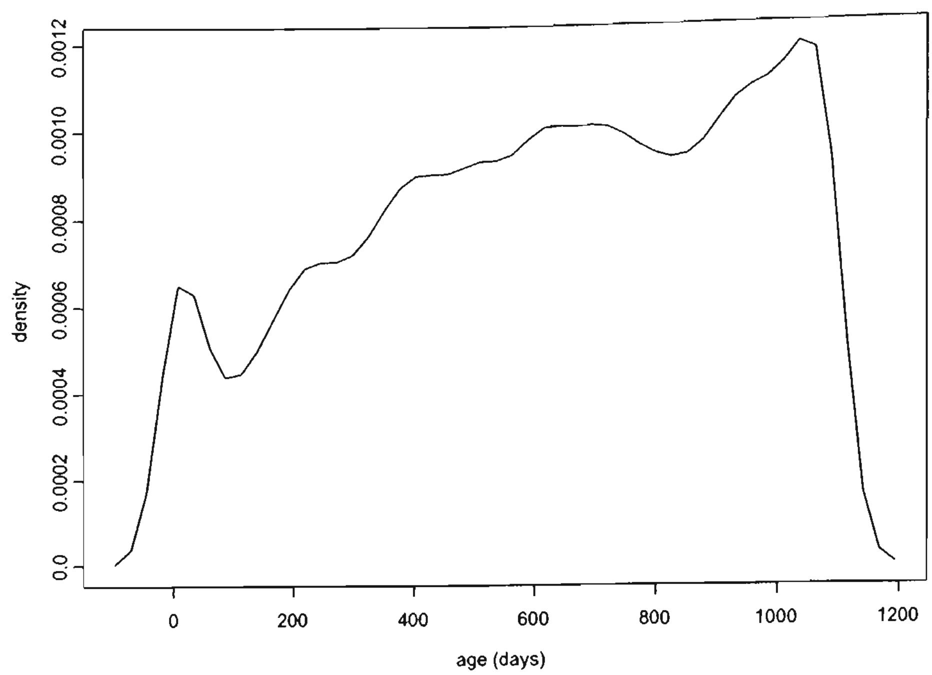


Fig. 3.2. *Claims density.*

On the other hand, a buyer may attempt to make a false claim, as may be the case if the item was misused. The manufacturer or repairer incurs a cost in processing the claim, and a further cost if the repair is carried out. If the manufacturer refuses to carry out the repair, further costs may be incurred in the form of litigation or the tarnishing of a reputation. The manufacturer then has to decide whether or not to honour the claim. As these issues are difficult to quantify, their inclusion adds complexities to the model. These complexities are not considered in the current study, and all claims in the warranty database have been assumed to be bona fide.

3.1.4 Modelling Rectification Action

A manufacturer usually has the option to repair or replace a product when it fails under warranty. If the product is expensive and complex, repair is usually the only viable option. The performance of the product after repair depends on the nature of the rectification action carried out. Murthy and Blischke (1992b) and Blischke and Murthy (1994) identified five main types of rectification actions:

1. *Replacement*: In the case of inexpensive items, or components of more complex ones, replacement is the usual option. The failure distribution of the new item or component would be the same as the original item, provided no design or production changes have taken place since the first item was produced.

2. *As-good-as-new repair*: Under this type of repair, a complex product is brought back to its original condition. The product's failure distribution after repair would be the same as that of a new item.

3. *Minimal repair*: Under this type of repair, the product is brought back to its condition just before failure. Its failure rate would be the same as it was just before repair.

4. *Between as-good-as-new and minimal repair*: The repair here may be a major overhaul, leaving the item with a lower failure rate than before the repair, but not as low as the failure rate of a new item. This may be modelled by a different failure distribution or distributions, which may depend on the number of repairs.

5. *Imperfect repair*: Under this type of repair, an item is left in a worse condition than before the repair. This could be due to a number of reasons, such as the installation of a faulty part, human error or the damage of another part during the repair process.

The effect that these repair-types have on the failure rate are represented in Figure 3.3, which has been adapted from Figure 2, Blischke and Murthy (1992b, p.5).

The analysis in this thesis is on a component level, with each repair consisting of a replacement of a component. Therefore, the new component has the same failure distribution as the original one, assuming that no design or manufacturing changes have taken place, and the rest of the components are taken to be unaffected. That is, the repair of the vehicle is taken to be minimal for all components other than the replaced one. Section 5.2 contains a more comprehensive discussion of the assumptions made in the modelling used in this study.

The time needed to carry out a repair consists of processing time, time to diagnose the problem, repair or replacement time, testing time, and time to return the item to the buyer, which can all be aggregated into one repair time. Repair time is important in warranties that include a penalty for down time, as is the case in a reliability improvement warranty (see Section 2.4). It is also of interest to buyers as this time deprives the buyer

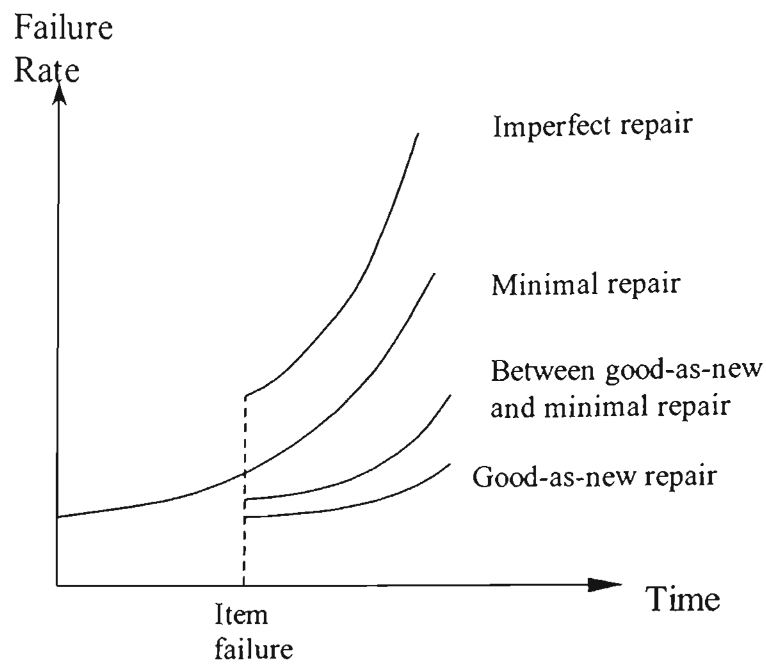


Fig. 3.3. *Failure rates after various repairs.*

of the use of the item, which may then lead to additional costs and loss of revenue. Repair time can be modelled as a random variable, having its own distribution. If the repair time is small compared to time between repairs, it can be taken to be zero. That is the case in this thesis.

3.1.5 Modelling Subsequent Failures

The modelling of time to failures after the first failure depends upon the rectification action after a failure, as discussed in the last section.

When a system is repaired as-good-as-new, the failure distribution of the repaired item is identical to that of the item when it was new. Similarly, a replacement part has the same distribution as the original if there have been no design or manufacturing changes. Both of these situations can be modelled by a renewal process, where the time between renewals has a distribution which is the same as the failure distribution of the first failure. A detailed formulation of the renewal process is shown in Section 5.6.

If an item is always subjected to minimal repair after failure, its failure rate after each repair would be identical to its failure rate just before failure. Thus, the item's first and subsequent failures can be modelled by a non-stationary Poisson process with intensity function given by the failure rate.

When an item is subjected to repairs other than as-good-as-new or minimal repair, the failure rate is not the same as that of a new item, or the item just before failure. Some works on imperfect repair appear in the literature, but these are from the inspection monitoring and maintenance viewpoint (for example, Pham and Wang, 1996).

3.1.6 Modelling the Cost of Rectification

When a warranty claim is made, the manufacturer incurs a variety of costs. These costs comprise an administrative cost (even if it is not a valid claim), the cost to repair or replace the item, a retailer's handling cost, a spare parts inventory cost and possibly a transportation cost to collect and return the item. These costs can be aggregated into one service cost. This service cost depends on the product characteristics and the usage patterns of consumers. Thus, it is a random variable which can be modelled by a suitable distribution function. If the variability in service cost is small, it can be treated approximately as a deterministic quantity. This seems to be the case in studies to date. The current study treats the cost of repair as a variable.

Let us first look at a model of the manufacturer's cost of servicing a free replacement warranty when a failed item is replaced. Suppose that c_s is the average manufacturer's cost of supplying an item. This cost includes all costs of doing business, including manufacturing, distributing and advertising, but excludes warranty costs. Let $C(t_w)$ be the total cost of supplying the item with a warranty of length t_w. Then the total expected cost of an item under warranty is

$$E\left[C(t_w)\right] = c_s[1 + M(t_w)], \tag{3.1}$$

where $M(\cdot)$ is the ordinary renewal function associated with the distribution function $F(\cdot)$ (see Subsection 3.4.5), and $E(\cdot)$ denotes the expected value. Renewal functions are difficult to evaluate in all but a few simple distributions, and this poses a problem for solving (3.1). However, extensive tables for various distributions are given in Baxter, Scheuer, Blischke and McConalogue (1981), and numerical methods are available, such as those provided by Xie (1989).

If the item is repaired instead of replaced, with the repair cost being constant (c_r), and the repair is as-good-as-new, then (3.1) can be modified to give the expected cost of supplying an item with a warranty of length t_w as

$$E[C(t_w)] = c_s + c_r M(t_w). \tag{3.2}$$

For minimal repair, between as-good-as-new and minimal repair, and imperfect repair, the $M(t_w)$ term in (3.2) needs to be replaced by the appropriate expected number of repairs with nonidentical failure distributions. Equations (3.1) and (3.2) can be found in a number of papers, such as Blischke and Scheuer (1975) and (1981).

For the data used in this thesis, repairs consist of replacing a failed component with a new one. As the analysis is conducted on a component level, the repair is taken to be as-good-as-new as far as the replaced component is concerned. The remaining components are taken to be unaffected by this replacement.

Cost models for the pro-rata, the combination, and the reliability improvement warranties, are given in Murthy and Blischke (1992b).

3.2 Modelling Two-Dimensional Warranties

The discussion in Section 3.1 applies to both one and two-dimensional warranties. Two-dimensional warranties, however, have the additional constraint of being limited by two variables instead of just one. For example, two-dimensional warranties exist for aeroplanes and automobiles. These warranties are limited by calendar time, taken from the purchase date, and usage, which may be measured as distance or flying hours.

There are two approaches to analysing a two-dimensional warranty. In the one-dimensional approach, analysis is performed on one dimension, with the second dimension being related back to the first dimension. This approach simplifies the problem to a one-dimensional problem. In the two-dimensional approach, analysis is performed using a two-dimensional distribution. This approach is mathematically more difficult.

Singpurwalla and Wilson (1993) stated that most of the work in modelling warranties has been restricted to the one-dimensional case, covering issues such as the expected cost of servicing a specific warranty, and the choice of an optimum price-warranty combination. They pointed out that not much has been done with two-dimensional warranties.

3.2.1 One-Dimensional Approach

In the first approach, item failure is modelled by an intensity function which is a function of both the age and usage of the item. The usage, in turn, is modelled as a function of age, so that the intensity function is expressed as a function of age only. This reduces the two-dimensional problem to a one-dimensional point process formulation.

Murthy and Blischke (1992b) and Murthy, Iskandar and Wilson (1995) formulated the one-dimensional approach as follows. Let the sale of the item occur at $t = 0$. Let $X_c(t)$ and $U_c(t)$ be the age and usage, respectively, of the item currently in use at time t. Let $U(t)$ be the total usage that a buyer has had from the purchased item, including any replacements, over the time interval $[0, t)$. If no replacements have occurred in $[0, t)$, then $X_c(t) = t$ and $U_c(t) = U(t)$. This is also true if the item is repaired minimally, and the repair time is negligible compared to time between repairs, and thus can be assumed to be zero.

If the relationship between age, $X_c(t)$, and usage, $U_c(t)$, can be assumed to be linear, then

$$U_c(t) = RX_c(t),$$

where R represents the usage rate which may vary from one user to the next. R can be modelled as a non-negative random variable with a distribution $G(r)$. Kim, Djamaludin and Murthy (2001) considered two different densities of $G(r)$: the uniform distribution and the gamma distribution. They also considered the discrete case where the usage rates can be categorised into groups. Thus, conditional on $R = r$, failures occur according to a Poisson process with intensity $\lambda(t|r)$. This can be modelled in the form

$$\lambda (t|r) = \psi (X_c(t), U_c(t))$$

where $\psi (x, y)$ is an increasing function of both x and y.

3.2.2 Two-Dimensional Approach

In the second approach, item failures are characterised by a two-dimensional distribution. Murthy, Iskandar and Wilson (1995) formulated the two-dimensional approach as follows. Let (T_1, U_1) denote the time and the item usage, respectively, at first failure. Similarly, let (T_i, U_i), $i > 1$ denote the time interval and item usage between the $(i - 1)$st and i-th failures. Then (T_i, U_i) can be modelled by a bivariate distribution function:

$$F_i(t, y) = \Pr(T_i \leqslant t, U_i \leqslant y)$$

The form of $F_i(t, y)$ depends on the nature of the rectification actions. For nonrepairable items, or if the repair is as-good-as-new, (T_i, U_i), $i \geq 1$, is a sequence of independent and identically distributed random variables with a common two-dimensional joint distribution function $F(t, y)$. For repairable cases other than as-good-as-new, the characterisation of (T_i, U_i) depends on the type of repair. An analysis of the failure data would be needed to determine an appropriate form of the distribution function, $F(t, y)$, which must have the property that $E[U|T = t]$ is an increasing function of t. Suitable distributions are the Beta Stacy and the Multivariate Pareto distributions. Iskandar (1991) provided a two-dimensional renewal function solver.

3.2.3 Comparison of the Two Approaches

Murthy and Blischke (1992b) pointed out that the one-dimensional approach requires the specification of the distribution rate, $G(r)$, across the buyer population, and the failure rate function, $\psi(\cdot, \cdot)$. If this can be done, the modelling can be done as a one-dimensional point process. This approach has the advantage that it is simpler to work with than the two-dimensional formulation. Many authors, for example Moskowitz and Chun (1994) and (1996), and Sarawgi and Kurtz (1995), have supported this approach, using a linear relationship between usage and time.

Murthy and Blischke (1992b) also pointed out that warranty data is generally collected in two-dimensional form. However, the manufacturer does not readily have any information after the warranty period, or about items without claims. Thus, an estimate for the usage rate, R, for each customer would have to be obtained from failure data, or from a post-sale survey.

It is anticipated that future work based on a two-dimensional distribution will be conducted.

In the current study, only the warranty claims data are available. It is shown in Section 5.2 that the data used in this study can be suitably modelled using a one dimensional approach. Since the analysis in this study is based on individual components, the one-dimensional approach is more practical because of the amount of analysis to be performed. It is anticipated that future work based on a two-dimensional distribution will be conducted. Gertsbakh and Kordonsky (1998) recognised that failure of mechanical components in automobiles, such as those in the braking system, predominantly depends on mileage. Others deteriorate over time due to environmental effects, such as humidity, as is the case with rusting of the car body. However, for vehicles without a warranty claim, only age and not distance is known, and usage would have to be estimated. This would introduce a source of error, so it has not been pursued. It would be desirable to further investigate this approach at a later date. Thus, the analysis of the data in the current study is one-dimensional, based on age.

Another reason for basing the warranty cost estimation on age, rather than distance, is the fact that many industries work on a monthly reporting cycle. As Elsayed (1996, page 513) pointed out: "It is more beneficial to the manufacturer to allocate warranty cost as a function of the age of the products, the number of claims during any time, and the number of products in service at that time." Thus, analysis based on age is more likely to be of use to the manufacturer.

3.2.4 Cost Models for Free Replacement Warranty

The cost models for the standard two-dimensional warranty bounded by a time t_w and a usage u_w, can be approached as a one-dimensional or a two-dimensional problem. Murthy and Blischke (1992b) give the sellers cost for nonrepairable items under a rectangular warranty bounded by a time t_w and a usage u_w as

$$E\left[C\left(t_w, u_w\right)\right] = \left[1 + M(t_w, u_w)\right] c_s,$$

where c_s is the cost to supply the item, and $M\left(t, x\right)$ is the two-dimensional renewal function associated with $F\left(t, y\right)$. This is the two-dimensional version of (3.1). The difficulty in solving this two-dimensional equation lies in evaluating the expected number of renewals, $M(t_w, u_w)$. The evaluation of $M(t_w, u_w)$ differs according to whether the one or two-dimensional approach is used.

If the item is repaired instead of replaced, with a repair cost of c_r, and the repair is as-good-as-new, the warranty cost under a rectangular warranty bounded by a time t_w and a usage u_w is

$$E\left[C\left(t_w, u_w\right)\right] = c_s + c_r.M(t_w, u_w),$$

This is the two-dimensional version of (3.2). Again, the difficulty lies in evaluating $M(t_w, u_w)$.

3.3 Modelling Product Reliability

A range of factors that influence how the reliability of a product is modelled has received attention by many authors, and is the subject of this section. The lifetime distribution can be affected by such things as the censoring time distribution and the delay in reporting warranty claims. Some authors have advocated the use of supplementary information to help model the reliability of a product more accurately. Many authors have included the influence of such factors by incorporating covariates into the reliability models. The following is a survey of literature that have discussed such issues.

3.3.1 Lifetime Distribution

Many authors have written about modelling the reliability of a nonrepairable product with a lifetime distribution. This is often a precursor to calculating the cost of a particular warranty policy. Lakey (1991) proposed a series of steps to determine the lifetime distribution, which include: the collection of data; the determination of the theoretical function that most closely resembles this density; and the estimation of parameters using maximum likelihood. These steps are broadly followed in the current study. She observed that within the defence industry there is very little expertise in assessing warranty risk for military weapon systems. The purpose of her work was to find the failure distributions which characterise selected families of equipment failures.

Ellis (1990) was motivated by the need to estimate the cost of extending a warranty on automatic transmissions used in medium and heavy duty on-highway vehicles and off-highway and military vehicles. He used data for major repairs only, which was defined as those repairs requiring the removal of the transmission from the vehicle. He proposed a two-component Weibull model, with a decreasing rate component, (Weibull shape parameter less than 1), and an increasing rate component (Weibull shape parameter greater than 1) to model the time to first failure. This caters to both infant mortality and wear-out with age. Similarly, Majeske and Herrin (1995) used a mixture of uniform and Weibull distributions to model time to first failure, suggesting an inherent defect component and a usage component, to model the time to first failure of two luxury-car components. Other studies have modelled the failure of a component on a combination or modification of distributions, such as a competing risk model involving two Weibull distributions (Jiang and Murthy, 1995), the beta-integrated distribution (Lai, Xie and Murthy, 1998), the exponentiated Weibull distribution (Jiang and Murthy, 1999), and the modified Weibull distribution (Lai, Xie and Murthy, 2003).

When a single component is to be anlaysed, these mixture models may be appropriate, especially if two sources of failure are evident. However, in this thesis, all the components of a vehicle are to be modelled. To make this task manageable, a single failure mode model is used.

Finding a method of selecting the member of a family of distributions that best fits a set of observations was the subject of Quesenberry and Kent's (1982) study. The distributions considered included the exponential, gamma, Weibull and lognormal. However, as the method requires a noncensored set of observations it is not used in this study.

3.3.2 Censoring Times

Some studies have investigated how the lifetime distribution of a product is affected by the distribution of censoring times, that is, the time when observation of a unit ceases. In warranty databases, only claims information is recorded. Nothing is known about units where there have been no claims. Suzuki (1985), Lawless, Hu and Cao (1995), and Hu, Lawless and Suzuki (1998) discussed the problem of using incomplete data to estimate the parameters of the lifetime distribution. They advocated the use of a follow-up survey to acquire usage data on units without a warranty claim in order to obtain better parameter estimates of the lifetime distribution. Similarly, Oh and Bai (2001) discussed the use of supplementary data that can be obtained from authorised dealerships who repair vehicles after the expiration of a warranty. The additional data, be it from customer surveys or from repair data, can be used as additional input in determining the parameters of the lifetime distribution.

Using failure-record data only, which is usually readily available, was an issue discussed by Kalbfleisch and Lawless (1988). The use of such data, they maintained, was insufficient to make satisfactory inference about the reliability of a product. They advocated the use of regressor variables determined by obtaining supplementary data, such as that obtainable through a post-warranty survey, or by using prior model data.

Although it is desirable to obtain additional information by surveying customers that have not made warranty claims, a survey was not possible in the current study as only the warranty data was available. Similarly, it was not possible to obtain post-warranty repair data from authorised dealerships.

3.3.3 Reporting Lags

Some authors, such as Kalbfleisch, Lawless and Robinson (1991), Lawless and Kalbfleisch (1992), Lawless and Nadeau (1995) and Lawless (1998) developed models that incorporated a time delay in entering a claim into the warranty database as a regression variable. They referred to this delay as the reporting lag. Kalbfleisch et al. (1991) found that a reporting lag of 20–60 days occurred in the data they used. In the warranty data that is used in this thesis, the data was entered into the database by the dealership on the day the customer brought the vehicle in for a repair. Additionally, sufficient time has been allowed for all claims data to be entered before the collection of the data for analysis. Thus, reporting lags are not considered in this study.

3.4 Survey of Approaches to Warranty Modelling

This section surveys the range of approaches to the modelling of product reliability and warranty cost that exists in the literature. Approaches have included the modelling of the number of warranty claims as a Poisson process; the Bayesian approach; the use of dynamic linear modelling (multivariate analysis); the Markovian approach; and lastly, the use of simulation. The use of the renewal function to model the time to failure is discussed in Section 5.6.

3.4.1 Warranty Claims Distribution

Some authors have modelled the number of warranty claims during a warranty by a non-homogeneous Poisson process. In this approach, the number of claims during the warranty period is of interest, not the type of repair or time to failure. Bohoris and Yun

(1995) used this approach to estimate the warranty cost under a hybrid (free replacement and pro-rata combination). Crow (1990) pointed out that complex systems such as automobiles, aircraft and communication systems, are repaired and not replaced upon failure. As such, the reliability of these repairable systems could be modelled as a nonhomogeneous Poisson process. Crow's model was a nonhomogeneous Poisson process with an intensity function

$$U(t) = \lambda\beta t^{\beta-1}, \quad t > 0,$$

with $\lambda, \beta > 0$. Because the intensity function has the same form as the hazard rate for a Weibull distribution, this particular nonhomogeneous Poisson process is often referred to as a Weibull Poisson process or the power law Poisson process. Crow (1993) went on to obtain confidence intervals for the failure intensity function.

Other authors have explored the use of different distributions to model the number of claims at a particular time. For example, Lu and Vance (1997) introduced the L-V distribution, defined as

$$\Pr(X) = \begin{cases} p, & x = 0 \\ (1-p)(1-r)^{x-1}, & x = 1, 2, 3, \ldots \end{cases}$$

with parameters p and r, where $0 < p < 1$ and $0 < r < 1$. They used the distribution to model the number of repairs per 100 vehicles. Lu and Vance used data from an authorised dealership containing all repairs of vehicles with nine or more months of service that were driven 10,000 or more miles. They found that the L-V distribution provided a better fit than the geometric distribution or the Poisson model. Lu and Vance's analysis provided a good snapshot of the situation at a particular time, but did not discuss the distribution of claims over time. The type of repair, the dispersion of costs of repairs and the usage rate of vehicles were not considered in their work.

Another issue that some authors raised was that claim rates may vary over time. Using the data from Kalbfleisch et al. (1991), Lawless (1995), Lawless and Nadeau (1995) and Lawless (1998) reported that by stratifying the data into six two-month production periods, it was evident that the claim rate was higher in one period than the other periods. Although there was some evidence in our data that the rate of claims does vary by month of production, it would be necessary to conduct that analysis on each component. This is beyond the scope of the present study, where an overall view is required. However, it would be worth pursuing in a future study, using a subset of components.

Lawless (1995) compared two methods for the analysis of recurrent events: marginal analysis (for example, rates and mean functions) and conditional specifications (for example, intensity functions) for covariates. He indicated that marginal analysis does not attempt to develop a full model for the event process, which could be used for predictions. The analysis in this dissertation attempts to model individual components. Thus, marginal analysis is not pursued in the current study.

The analysis in this thesis is done at a component level. Thus, this thesis is not concerned with the aggregate number of repairs at any particular time on the system as a whole, but with the reliability of individual components and the overall warranty cost due to these claims. As such, the nonhomogeneous Poisson process is not used to model the reliability of vehicles; a renewal function approach on individual components is used instead.

3.4.2 Bayesian Approach

In the Bayesian approach to reliability analysis, supplementary information is used to establish a *prior* probability density function, $g(\theta)$, where the form of the distribution is known or assumed and the value of the parameter θ is sought. "This distribution expresses the state of knowledge or ignorance about θ before the sample data are analysed." (Martz and Waller, 1991, page 167.) The supplementary information may come from engineering design and test data, operating data in different environments, engineering or expert judgement or even operating experiences with similar equipment. Using the prior distribution and the operating data y (for example, warranty data), the *posterior* probability density function, $g(\theta|y)$ is established.

A number of studies have used the Bayesian approach, for example, Giuntini and Giuntini (1993). In this thesis, the prior probability density is established for individual components. A posterior probability density function is not sought, as all the available data is used to establish the prior density function. However, new data can be used to establish a posterior density in future studies, once the prior densities have been determined.

3.4.3 Dynamic Linear Modelling and Neural Networks

In forecasting claims when a new model is introduced, Wasserman (1992) recommended the inclusion of prior model data, if there have not been significant changes in the manufacturing process or warranty policy. Wasserman and Sudjianto (1996) compared three strategies for forecasting warranty claims. The first, the static predictive model, involves the use of either regression or time series models. They are the most widely used. They are adequate for interpolation, but are risky when extrapolating using data that is highly non-linear, or that is perturbed by localised phenomena. The second strategy is the Kalman filter or dynamic linear modelling strategy, which has the ability to adapt to local process trends while maintaining the simplicity of a linear model. The third is the nonparametric modelling approach, which provides the flexibility to capture nonlinearities in a data set. However, if not fitted carefully, the model may be too responsive to local variations. This approach includes the use of artificial neural networks. Wasserman and Sudjianto show that the nonparametric modelling approach provided a better fit to the test data than did the use of the other two strategies. This is because it offered the flexibility to model local trends and any general nonlinear phenomena.

The appeal of neural networks is that they are relatively easy to program, there is no requirement to use any particular equational form, they appear to be highly resistant to outliers and commercial packages are widely available (Stern, 1996). One disadvantage of neural networks is that they do not give prediction intervals indicating the level of uncertainty. Stern suggested that the neural networks approach is appropriate when the phenomenon is complex, involving many predictor variables in a nonlinear combination, and when a sufficient amount of data is available to estimate a complex model. However, if the underlying process is well understood then it is usually advantageous to develop a model.

Since we have established mathematical models to describe the warranty data, and we are interested in predicting costs under an extended warranty, the use of dynamic linear modelling and neural networks does not appear to be the best approach in this situation.

3.4.4 Markovian Approach

As Trivedi (1982, page 309) stated, a stochastic process is a Markov process if the probability of an event occurring depends only on the state of the system at the present time, and not on its past history. Thus, the distribution that describes the probability of an event occurring in the future depends on the present state of the system, and not on how the system arrived in that state.

Some products, such as clocks, refrigerators and industrial pumps, are used continuously, whilst others, such as television sets, washing machines and automobiles, are used intermittently. The failure rate during usage would be different from that when idle. To obtain the distribution function for the time to failure, it is necessary to model the usage pattern. Murthy (1992) modelled the transitions between in-use and idle periods as a two-state continuous time Markov process. He examined what happened when the usage intensity varied, and found that the cost of a warranty increased with increasing usage. In this thesis, we are not concerned as to whether failure occurs whilst a vehicle is idle or in use. In order to keep the modelling manageable, the reliability of components over time is modelled without concern about the fact that usage is intermittent.

A different approach using a Markov process, is one where the reliability of an item is modelled as a transition through compartments that describe the aging process. The probability of the transition from one state to a different state can be calculated from the probability of failure of the item. This approach was used by Faddy (1993). In yet another Markovian approach, Balachandran, Maschmeyer and Livingstone (1981), considered a three-component system. The item went from one state of component condition (working/failed) to another state. The probability of component failure can be used to work out the probability of transition from one state to the next. Clearly, when a product comprises many components, as in our case, this method becomes very cumbersome. Blischke (1990) commented on the problematic nature of this approach, even though the transition probabilities are more easily obtained than the renewal functions. Thus, this approach is not used in this study.

3.4.5 Renewal Process Approach

The replacement of a component with a new one, or a repair that is as-good-as-new, can be modelled by a renewal process. The renewal function, $M(t)$, counts the number of replacements that occur in a given period of time. Frees (1986) developed three estimators of $M(t)$ which could be used to estimate the warranty cost of a product. Frees and Nam (1988) used a straight line approximation to evaluate the renewal function under two different warranty policies made up of combinations of free replacement, pro-rata and renewing warranties. They found that this method gave reasonable results. Frees (1995) went on to develop a semiparametric estimation of warranty costs. An example of the application of the renewal function to estimate warranty costs can be found in Summit and Cerone (2003), where this approach was used to analyse the failure data of an automobile component.

This current thesis uses a renewal process approach. The formulation of the renewal function is discussed in more detail in Section 5.6.

3.4.6 Simulation Approach

Many papers have been written with the aim of offering support to management by providing a software tool that can be used as an aid to decision-making, for example, Rooker (1989), Isaacson and Brennan (1991) and Hadel and Lakey (1993). Monte Carlo simulation is used to generate failure times and warranty costs. Typically, inputs include mean time between failures, the form of the failure distribution and its parameters, the length of the warranty period, the average cost of repair, the form of the rebate function, and the number of replications of the simulation required. In some of these studies, the model used to generate results was not detailed.

In contrast, Hill, Beall and Blischke (1991) presented a simulation where one of three lifetime distribution could be chosen: the gamma, Weibull or truncated normal distributions. Inputs to the simulation included the parameter values of the lifetime distribution, the warranty length, the length of the life cycle of the product, the form of the rebate and discounting functions. The simulation yielded the mean and standard deviation of the set of simulated costs. Hill et al. validated the model against the true expected costs given by the mathematical model in the exponential case. They found that the warranty length, rebate terms, and the lifetime distribution of the product were the main determinants of warranty cost.

In a study by Roberts and Mann (1993), the failure of major components of a system was simulated. They found that the simulation supported the nonhomogeneous Poisson process model (Crow, 1990). The simulated model had the advantage of identifying which major component had failed, thus giving more information than the nonhomogeneous Poisson process model. In this thesis we extend the simulation to all components and use simulation to verify the results of the theoretical model.

In their synthesis of existing warranty models, Hill and Blischke (1987) concluded:

It would appear that the "real-world" manager seeking a working model for quan-

tifying the cost of a given set of warranty provisions may not find a suitable one among the published mathematical models. . . . the existing models are capable of yielding a closed-form solution to the expected cost of a warranty when the life distribution is assumed to be exponential . . . the authors suggest the use of a computer simulation model as an alternative flexible enough to incorporate virtually any type of product failure distribution.

In this thesis, simulation is used to validate an overall warranty cost model based on the Weibull and exponential reliability models of individual components within a vehicle.

3.5 Conclusion

The purpose of this chapter has been to discuss the modelling of the warranty process, the approaches to the modelling of two-dimensional warranties, various aspects of the warranty-modelling process that some studies have explored and the various approaches to warranty-cost modelling that have appeared in the literature. The approach that is used in this thesis is the renewal process approach, which is developed in Section 5.6.

The next chapter discusses the manufacturer's data. It looks at the processes involved in getting the data to a usable state, including importing the data into S-Plus, exploring the data and building up the survival database. The theoretical models are discussed in Chapter 5, including the renewal process approach to modelling the number of claims during the warranty period. This is followed by the application of the models to the cleaned manufacturer's data in Chapter 6.

Chapter 4
The Manufacturer's Data

4.1 Introduction

This chapter details the data that are used in this study. The source of the data, the data's shortcomings and how they have been overcome are discussed in this chapter. The detail of the procedure involved in going from raw data to a database that is ready for survival analysis is outlined.

As established in Chapter 1, published warranty data, in particular Australian data, is limited. This study analyses some genuine warranty data from an Australian automobile manufacturer. To protect from misuse of the data, the manufacturer's identity has been withheld. The manufacturer has made available the production details of vehicles manufactured during 1997, including all the warranty claims on these vehicles right up to the download date, 21/5/2002. The raw data indicates that the 3-year warranty had expired on 99.95% of the vehicles in the database.

The data that are analysed in this study comprise two databases. The first contains 30,117 records of vehicles that were manufactured during the year under study. The fields in the sales database include vehicle identification number (VIN) and production date (*prod*). Other fields are included in the database but have not been used in the analysis in this study.

The second database contains 62,456 records of warranty claims made on those vehicles. This database of claims has been built up by the dealerships, who, on behalf of the manufacturer, have fulfilled the warranty claims made by customers. The fields in this database include vehicle identification number (VIN), dealer identification ($dealer$), claim number ($claim$), production date (*prod*), sale date ($sale$), repair date ($repr$), distance ($dist$) in kilometres and age at claim (age) in months, replaced part number ($part$) and name ($partname$), and cost ($cost$) of repair. There are other fields included in the database that have not been used in the analysis.

S-Plus 2000 as been used extensively in the analysis of the data. It has been chosen for this project because S-Plus is a powerful data analysis product that has its own built-in

S language. This allows for the writing of scripts, thus providing a means of re-using programs with different data sets. It is an aim of this study to produce reusable code, so that the analysis of other data sets can be carried out. S-Plus's power lies in the fact that it works on vectors and arrays of variables, rather than on the processing of individual values within a vector. This was an important consideration because of the large size of the databases. In addition, S-Plus offers a comprehensive number of functions that can be used in survival analysis. Apart from S-Plus, packages that were considered include SPSS, SAS, Statistica and Minitab.

A number of original S-Plus scripts and functions have been written by the author of this thesis, in order to conduct the analysis described in this chapter. Many are presented in Appendix A. All the functions, scripts and data frames referred to in this chapter can be viewed on the accompanying compact disc.

The rest of this chapter describes in detail what has been done with the data. As this is a practical project using real data, considerable time and effort have been invested in exploring the data, validating it and cleaning it. The chapter starts with a discussion of the format of the downloaded data, and how it has been imported into S-Plus. The nature of the exploratory data analysis is discussed in the subsequent section. The way in which the errors in the data have been identified and treated is discussed extensively in the section following that. The treatment of the data pertaining to vehicles that have been repaired more than once is then discussed. Subsequent to this is a section that discusses the building up of the database that has been used to conduct survival analysis. The chapter concludes with a summary of the data cleaning that has been carried out in this study.

4.2 Importing The Data

The manufacturer's data consists of three Excel files. The first contains sales information, and the other two contain a claims database which, due to its size, had been split into two for downloading purposes. All three files have been converted to tab-delimited ASCII files firstly, then imported into S-Plus. By converting the Excel files to tab-delimited files, it has been possible to set field data-types for the import. Before converting the two claims files, the field *id* has been added to each database. Filling this field with the concatenation of VIN, *dealer* and *claim* has created a unique key to the records, enabling duplicates from the two separate files to be identified.

To import the sales data, the script *ImportSale.ssc* has been written, and can be found on the accompanying compact disc. The script adds the field *age*2, which is to hold the age of vehicles at the time of a claim. This field has been included in the sales database in order to have an age for vehicles that have had no claims. Its value is the number of days between 21/5/2002 (the file download date), and the vehicle production date, less 28 days. The 28 days represents the average delay between production and sale of a vehicle (see Subsection 4.4.2). The value of *age*2 has been limited to three years, this being the time limit on the warranty, after which observations are censored. In the claims database, *age*2 has been calculated from the sale and repair dates primarily, with the fields *age* and *dist* also available. When *age*2 could not be calculated from the data in the claims database, it could be estimated using the value in the sales database. Obtaining values for *age*2 is discussed further in Section 4.4. Thus, the script *ImportSale.ssc* creates a database of vehicles produced in 1997, including an estimate of their age on 21/5/2002. This sales database is the S-Plus data frame *sales97* on the accompanying compact disc.

The script *ImportClaim.ssc*, which is included on the accompanying compact disc, imports the two claims data files. The field *age*2 is added after the import, and holds the difference between the repair and sale dates. This field is used to calculate the vehicles' ages at the time of a claim. Then the script combines the two claims files into one, eliminating any duplicate records by using the field *id*, as described above. The next section discusses how, in order to check for errors, *age*2 is compared to the field *age*, which is included in the manufacturer's database.

This claims database needs to be sorted by VIN, then by *part* and then by *repr*, to facilitate calculating a component's age, as discussed in Section 4.5. Because of the size of the database, the sort takes a long time in S-Plus, but when exported to Excel, it can be done quickly. It is possible to sort all the data in Excel initially, before converting the file to a tab-delimited format, if there is only a single claims file. However, in the case of two claims files, the data needs to be sorted after the two claims files are combined. Once the data is sorted, vehicles that have been repaired more than once can be identified, and a component's age within these vehicles can be calculated. After being sorted, the Excel file has been converted to a tab-delimited ASCII file so that it can once again be imported into S-Plus in the appropriate format. A field to keep track of any data changes, and another field to flag errors have been added. The claims database is the S-Plus data frame *claims97* on the accompanying compact disc.

The original manufacturer's data files have not been included on the accompanying compact disc. They are omitted so that the identity of the manufacturer cannot be deduced from the original data. Some fields have been omitted from the original data and the part numbers have been changed before the import. However, the imported data is available on the disc and can be viewed in an S-Plus session.

4.3 Exploratory Data Analysis

As an initial step, the data have been explored in the manner described in Tukey (1977, Chapter 2), in order to obtain an overall impression and to identify any patterns of errors present. Boxplots of all the numeric variables, *prod, sale, repr, age, age*2, and *cost*, have been plotted to find any outliers in the data. As can be seen in Figure 4.1, all of these variables except *prod* contain outliers. A matrix plot of the three date fields has been constructed to examine the relationship between these fields. However, little can be concluded from this plot due to the existence of extreme outliers.

Summary statistics, including minimum, maximum, upper and lower quartiles, median, mean, standard deviation and the number of records containing NA (not available) have been generated. These statistics, together with the plots outlined above, are a very useful starting point in identifying errors in the data. An S-Plus function called *fPlotSmr* has been written to generate the plots and summary statistics. The function is presented in Appendix A.1.

It is evident from the exploratory data analysis that the data contains a number of errors, and that further investigation is needed to identify the types of errors, the number of cases of each error type, and the actual cases in error. To facilitate this process, an S-Plus function *fCheck* has been written. Figure 4.2 shows the process flow of the function and the complete function is shown in Appendix A.2. The output of this function is also displayed in Appendix A.3.

A discussion of the anomalies that have been found in the data, and an indication of the number of records affected, follows. Most of the figures below have been obtained

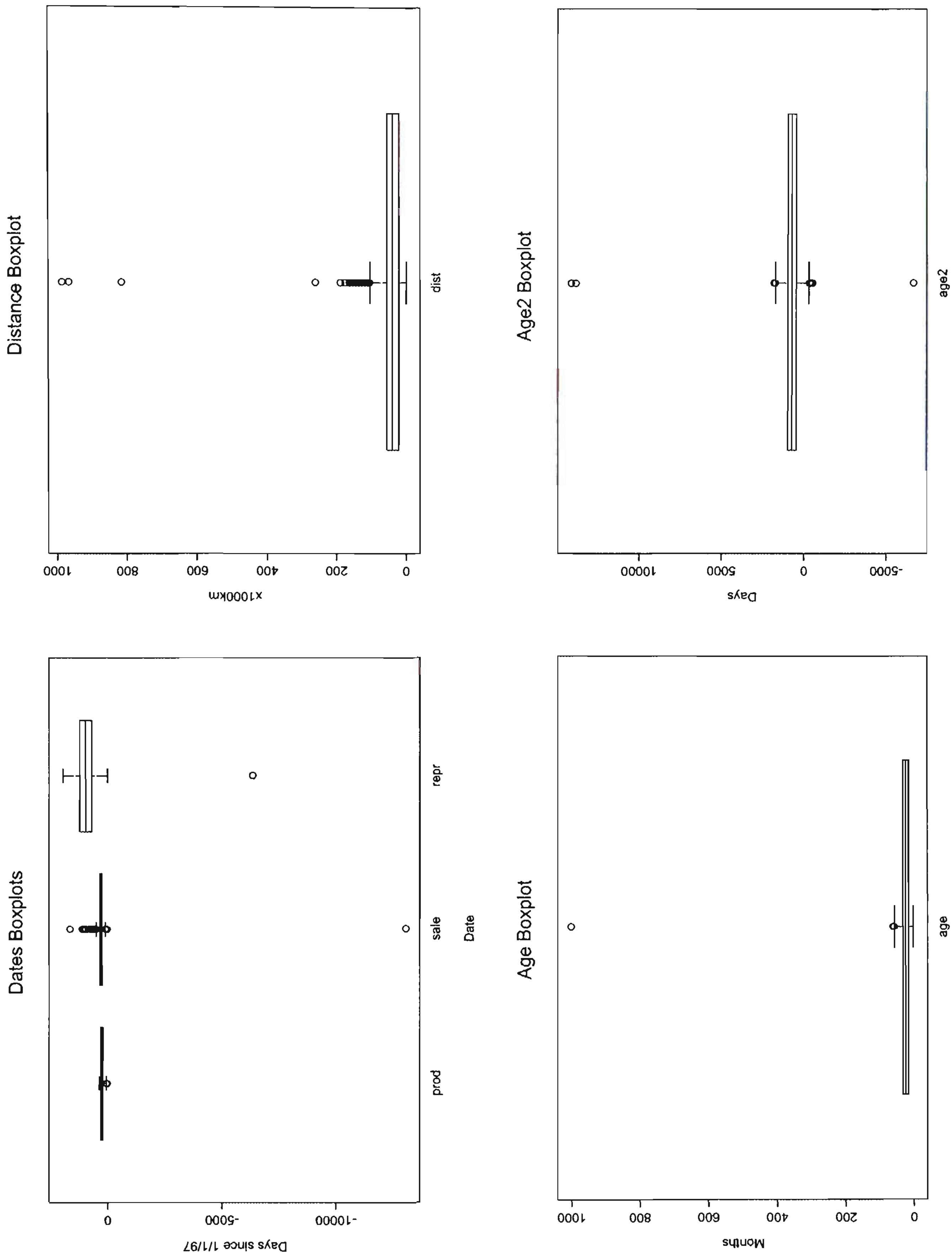


Fig. 4.1. *Raw data box plots.*

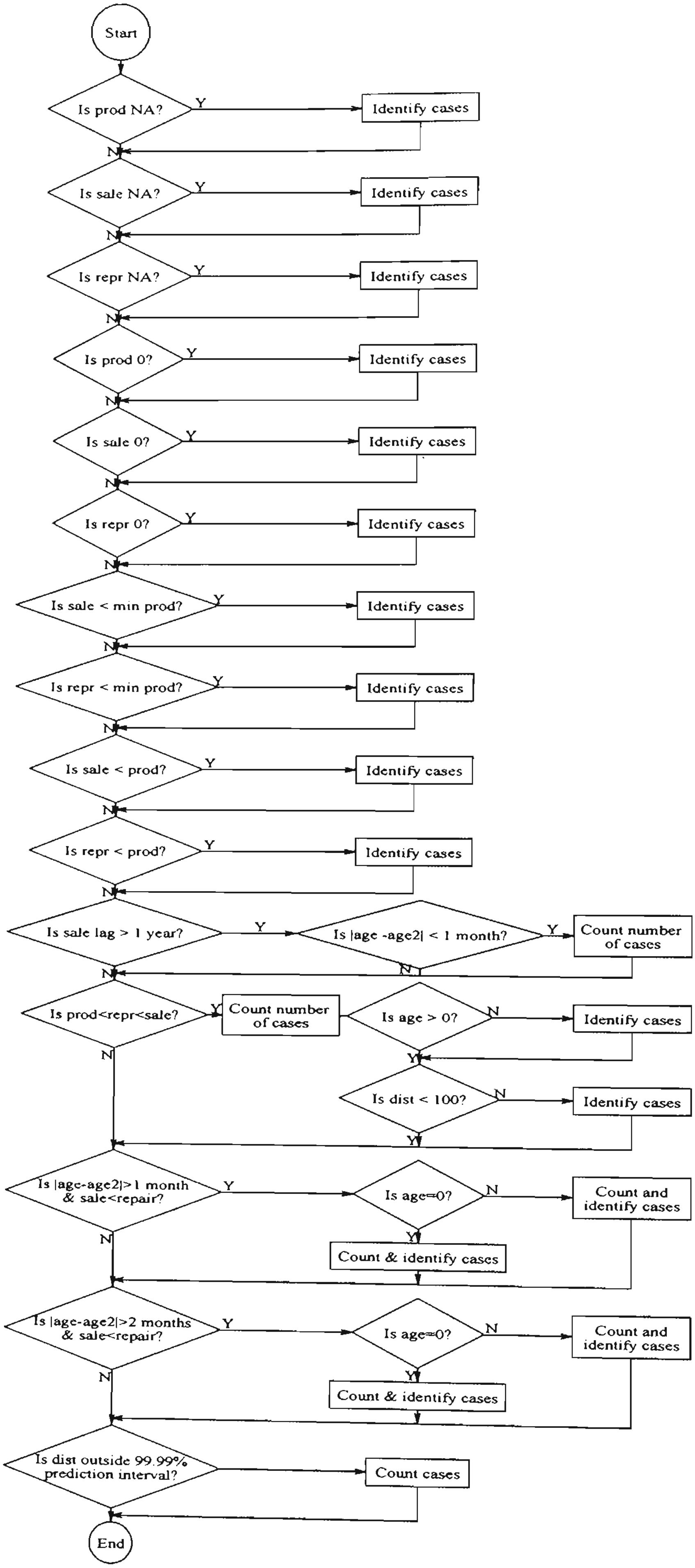


Fig. 4.2. *Logic flow for data checking function.*

from the output of the function *fCheck*, whilst others have been derived separately. The section following this contains a discussion of how these problems have been resolved.

1. *Sale and repair dates are either not available or zero:* The sale date is NA in 8 records. All records contain a repair date, and none of the sale or repair dates contain zero. (This is not the case for other years' databases.) The manufacturer's representative has indicated that an NA (or zero) in the sale date means that the vehicle was repaired before it was sold, thus making a sale date unavailable.

2. *Sale date is before production date:* There are 5 records. This is obviously impossible and must be due to an incorrect data entry. However, in all of these five records, the sale date is after the first production date (1/1/97), so the error may be in the production date. Therefore, these production dates need to be checked against the sales database.

3. *Repair date is before production date:* There are 17 records. Again this is impossible, and these records must be incorrect. The repair dates are not before 1/1/97, so the production dates may be in error in these records, and therefore need to be checked.

4. *The sequence of events is production, then repair, then sale:* There are 753 cases of this situation. At first glance, a warranty claim before the sale of a vehicle seems impossible. But this could again have been a case of a vehicle being repaired before it was sold. This could have been initiated by the dealer upon delivery of the vehicle. Examples of such repairs include the removal of marks on duco, upholstery or interior coverings, the replacement or recharge of a battery, the lubrication of squeaky hinges, the replacement of missing or damaged trim or even mechanical adjustments such as a front-end alignment. Any of these items could possibly have been picked up by a buyer upon the inspection or test drive of a vehicle. This conjecture is supported by the fact that in 365 of the 753 cases, the sale occurred within two weeks of repair.

5. *Sale date is unrealistically too long after production date:* In 120 records, the sale date is more than one year after the production date, with 21 of these being over two years after the production date. The median number of days that sale occurs after production is 28, and the maximum is 1357 days, the latter clearly indicating an

outlier. The data appear to contain genuine cases of long delays in selling vehicles, as well as errors which need to be cleaned.

6. *Age does not equal the difference between repair and sale dates:* In 111 records, the difference between age and $age2$ is greater than one month. This figure excludes the cases where repair is before sale. Of these, 77 have over two months difference between age and $age2$. Although the manufacturer's representative stated that checks have been put in place at the data entry level for the age field, this clearly could not have been the case. In 16 of these 111 records, the value entered for age is zero, and in many cases, a zero in the age field does not match the distance entered. For example, in 89 cases, $dist$ is over 100 kilometres whilst age is 0. This suggests that perhaps some dealers may have entered zero in the age field simply for expedience.

7. *Distance does not correspond to the age of the vehicle at repair:* To check the value entered in the distance field of a record, it has been compared to the 99% prediction interval for distance based on two linear models, each going through the origin. The first model uses age as the independent variable, and the second uses $age2$. Thus, the models are of the form $dist = \beta_1 age + \epsilon$, where$\beta_1$ is a constant and ϵ is the error term. Bowerman and O'Connell (1990, page 193) give the $100\,(1-2\alpha)\,\%$ prediction interval for the linear model through the origin, $y_0 = \beta_1 x_0 + \epsilon$, as

$$\left[\hat{y}_0 \pm t_{(\alpha,n-1)} s \sqrt{1 + \frac{x_0^2}{\sum_{i=1}^{n} x_i^2}}\right], \tag{4.1}$$

where $t_{(\alpha,n-1)}$ is the α percentile of the t-distribution with $n-1$ degrees of freedom, and s, the sample standard deviation, is given by

$$s = \frac{\sum (y - \hat{y})^2}{n-1}.$$

The data contains 585 cases where $dist$ is outside the 99% prediction interval using age as the independent variable. Similarly, there are 715 cases outside the 99% prediction interval using $age2$. Thus, errors amongst the fields $dist$, age and $age2$ do exist in the data.

8. *Extreme values in the cost of repair*: The two highest costs are \$23,636.30 and \$27,193.00. It appears that these values could be in error. However, the manufacturer's representative has indicated that such extreme cases, although rare,

have actually been paid out. Therefore, these costs have not been cleaned out of the database.

It should be noted that these error groups are not mutually exclusive. That is, some records contain more than one anomaly. Overall, these anomalies represent less than two percent of the claims data, making the data reasonably accurate. Thus, removing records with these errors could have been an option, but so doing would have introduced a bias in the number of failures in the subsequent survival analysis. The next section discusses the approach that has been taken in handling these anomalies, and explores more deeply the criteria that have been used to determine whether an anomaly is in fact an error.

4.4 Data Cleaning

Blischke (1990) was aware of the problems of inaccurate data when he wrote:

> *User data are frequently only haphazardly obtained and can be notoriously unreliable. Some models take this into account by including parameters which express probabilities of false warranty claims and of undeclared legitimate claims. How one might estimate such parameters is another unresolved problem.*

Rai and Singh (2003) discussed the problem of unclean warranty data. In finding the distribution of usage rates, their approach was to omit data that was outside the 99% statistical limits of the data.

The concern of this study is to obtain an accurate set of data that can be used to obtain models for the reliability of components. Attempting to factor in data inaccuracies into the models can make them very complex and unusable. The approach that has been taken in this study is to clean the data as far as possible.

As stated in Subsection 3.2.3, the analysis in this study is based on age, rather than on usage. Thus, it is important to accurately establish components' ages from the fields age, $dist$ and $age2$. The fields age, $dist$, $repr$ and $sale$ have been entered by the dealerships, thus making them susceptible to entry errors, and therefore in need of careful checking. Since some obvious errors exist in the age field, it has been decided that $age2$ should be used as a starting value for the vehicle's age, as the sale and repair dates are less likely to be in error. The value in $age2$ can then be checked against the other fields.

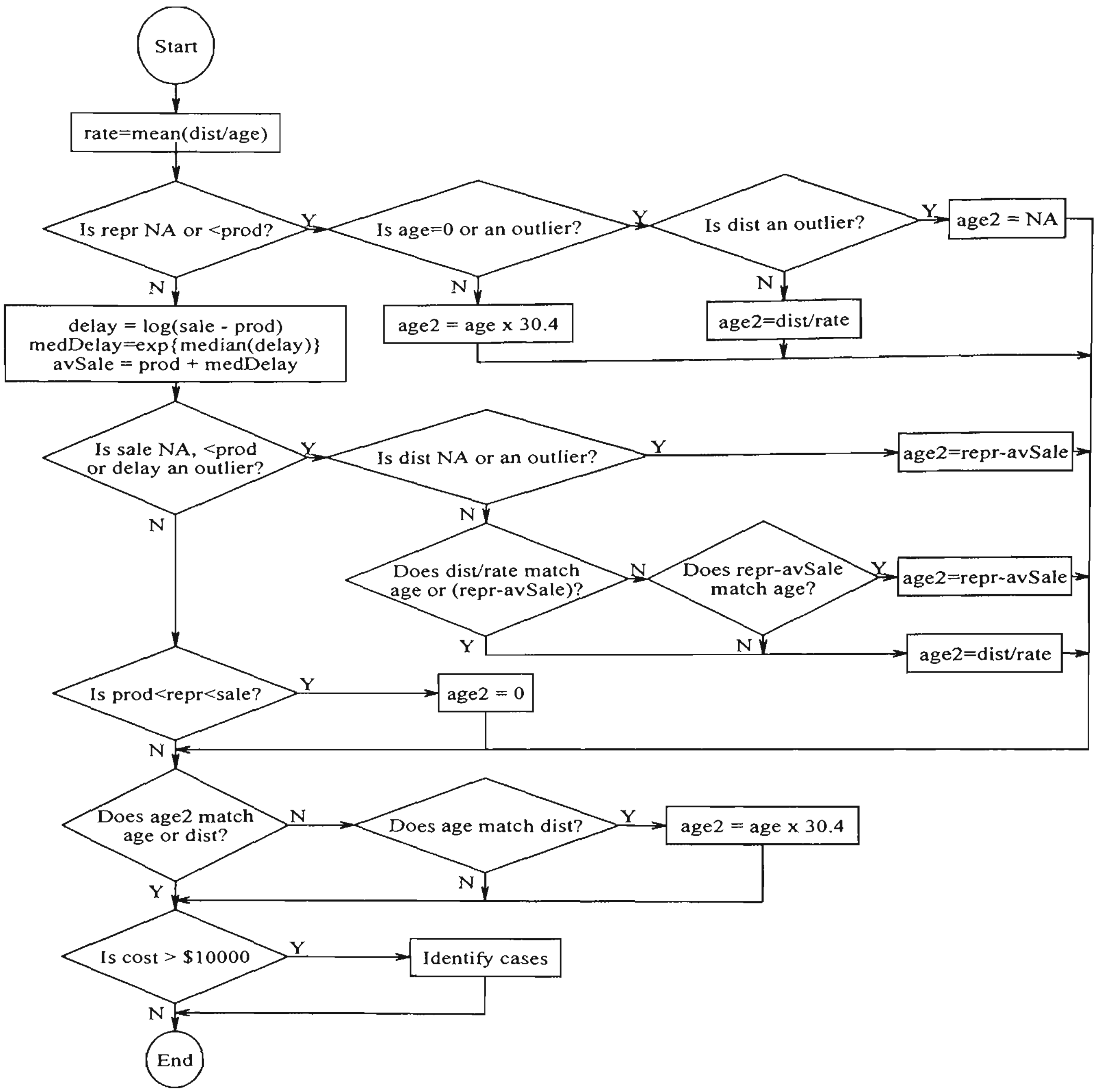


Fig. 4.3. *Logic flow of cleaning function.*

A starting point for checking for errors in *age*2 is to look for errors in *repr* or *sale*, since *age*2 is based on these two fields. After that, *age*2 should be checked against the values of *age* and *dist*, which could be used if the value in *age*2 is found to be in error.

An S-Plus function, *fClean*, which is presented in Appendix A.4, has been written to clean the data. A flowchart showing the logic of the function is shown in Figure 4.3. A discussion detailing the data cleaning procedure that has been undertaken follows in the next section.

4.4.1 Errors In The Repair Date

It can be seen from Figure 4.3 that the first error in $age2$ to be dealt with is the one originating from an error in *repr*. Errors in the repair date can be classified into three types: (i) value is NA, (ii) repair is before production, or (iii) repair is before the sale, but after production. This last case is discussed in Subsection 4.4.3 below. To clean records containing the first two types of error in the field *repr*, the fields age or $dist$ are available to establish the vehicle's age. If age is not 0 or an outlier, then $age2$ has been given the value of age multiplied by the average number of days in a month. Otherwise, if age is 0 or an outlier, and if $dist$ is not an outlier, then $age2$ has been set to $dist/rate$, where $rate$ refers to the usage rate, measured in kilometres per day. To reduce the influence of outliers, $rate$ has been obtained using the S-Plus function *lmRobMM*, which fits the robust and efficient MM-estimator proposed by Yohai, Stahel and Zamar (1991). A zero intercept has been imposed on the linear model, and $rate$ has been assigned the value of the model's gradient. If neither age nor $dist$ are usable, then $age2$ needs to be set to NA in the function *fClean*. However, no records in the data used in this study have needed to have their value of $age2$ set to NA.

4.4.2 Errors In Sale Date

Before *sale* is examined for errors, the field *prod* needs to be checked for errors. This is done because *prod* is needed to check the value of *sale*, and in some circumstances, is needed to estimate the sale date. The values entered in *prod* in the claims database are easily checked against the values entered in the sales database for each vehicle, by matching records on VIN. All values in *prod* were found to be correct, and there are no blank or NA entries.

The next field to be checked is *sale*, and as is the case for *repr*, there is a need to search for zeros, NA and outliers. In addition, records for which the sale date is before the production date, or for which the sale date is unrealistically too long after the production date, need to be found. The *fCheck* function described in the last section identifies records with all of these problems with the exception of the outliers in *sale*, so this has been tackled next. Since the sale dates are closely linked to the production dates, which span a year, the range of sale dates is closely packed with a range not much wider than one year. Perfectly acceptable sale dates at the beginning and end of the year have been found to be outliers. Thus, it is not practical to exclude all outliers in $sale$. The approach that has been adopted is to find extraneous values in $sale$ that have a sale date an uncharacteristically long time after the production date, which has already been checked.

We define $delay = sale - prod$. After removing records in which $sale$ is 0, NA or less than $prod$, the distribution of $delay$ has been examined. It was found to be skewed, with a minimum value of 0 and a maximum of 841 days. Because of the skewed nature of $delay$, the outlier limit on its natural logarithm scale has been used as a cut-off value. This worked out to be 372 days, which is a fairly long time (just over one year) for a vehicle to remain unsold. However, its value is plausible, so it has been accepted. In records where $delay$ is more than 372 days, the value in $sale$ has been taken to be an error, and the value of $age2$ has had to be recalculated, as described below. This affected 83 records.

We defined $medDelay$ as the median of $delay$, which is 28 days. A representative sale date could be obtained by adding $medDelay$ to $prod$, which could be subtracted from $repr$ to obtain an approximate age for a vehicle. However, in cases where the repair date is before this calculated representative sale date, the vehicle's age would have to be set to 0. That is, $\max\left[0, repr - (prod + delay)\right]$ could be used as an approximation to the vehicle's age. As this value is an approximation to the vehicle's age, it would only be used if nothing else was available.

Returning to the logic flow diagram of Figure 4.3, it can be seen that in checking the value of $age2$, the function *fClean* checks the value of $repr$ first, and then determines if there are any errors in $sale$. Once it is determined that there is an error in either of these fields, the function uses $dist$ as first choice, then age and then $\max\left[0, repr - (prod + delay)\right]$ to evaluate $age2$. In any event, a match is sought between this new value of $age2$ and one of the remaining two values. The complete procedure of the function *fClean* can be seen in Figure 4.3.

4.4.3 Repair Date Between Production And Sale

As $age2$ has been set to $repr - sale$, the value in $age2$ would be negative if the repair date for a record is before the sale date. Obviously a vehicle cannot have a negative age, so these records need to have their values of $age2$ recalculated.

As discussed in the previous section, a vehicle could have been repaired before it was sold. Thus, a repair date prior to a sale date is not necessarily an indication of an error, provided the repair and sale dates occur after the production date. There are 753 records satisfying these conditions, seven of which have already been dealt with because their sale date is too long after production. This leaves 746 records that need to be assessed. In 743 of these, *age* is zero, in two records *age* is one month, and in one record *age* is two months. In the latter two cases, *age* could have been calculated from the production date by the dealerships. One could conclude from this that the vehicles had most likely been repaired before they were sold. However, in only 712 of the 746 cases *dist* is less than or equal to 100 kilometres, and in a further 14 records, *dist* is less than 200 kilometres. It could therefore be concluded that because of the relatively small distances travelled by these 712 (perhaps even 746) vehicles, a repair was carried out before sale. But in the remaining records, *dist* is significantly larger, up to 99,565 kilometres in one case. It was unclear as to where the error lies in the eighteen cases where *age* is two or less months and *dist* is over 200km. Either there is an error in the value entered for *dist* or there is an error in the values entered in *age* and *repr*. It is more likely that there is an error in one field (*dist*) than in two fields (*age* and *repr*). Thus, *age*2 has been set to zero in all 746 records.

4.4.4 Check *age*2 Against *age* And *dist*

After all of the above processing, the value of *age*2 has been verified against the values of *age* and *dist*. This is done within the *fClean* function. Despite the redundancies in checking the adjusted *age*2 values, which have previously been verified as described above, it is more efficient to check all the values of *age*2, as S-Plus works on the whole vector of values altogether. If the value of *age*2 cannot be verified against either *age* or *dist*, then the value of *age* is checked against the value of *dist*, and if these values correspond to each other, *age*2 is given a value based on *age*.

The value of *age*2 is deemed to match the value of *age* if their values are within one month. Since *age* is measured in months and *age*2 is calculated in days, a difference of up to one month in their values is deemed to have verified the value in *age*2.

The fields *age*2 and *dist* are considered to be in agreement if the value of *dist* lies within the (two-sided) 99% prediction interval of distance for the given value of *age*2. The prediction interval is given by (4.1), with the dependent variable, *dist*, being linked to the independent variable, *age*2, by the linear model described in Subsection 4.4.1.

If $age2$ does not match either age or $dist$, but the value of $dist$ lies within the 99% prediction interval for the given value of age, then $age2$ is assigned a value of $age \times 365.25/12$ in the function *fClean.*

In the data used in this study, there are seven cases where the values of either age or $dist$ do not lie within the 99% prediction interval for the given value of $age2$. In only two of these does the value of $dist$ lie within the 99% prediction interval for the value of age. In these two records $age2$ has been assigned the value of $age \times 365.25/12$. In the other five cases, the value of $age2$ has been left as is.

4.4.5 Cleaned Database

All of the cleaning and checking of data described above have been incorporated into the function *fClean*. The result of passing the claims database to the function is a cleaned database, which is the data frame *clean97*. The file for this data frame is on the accompanying compact disc, and it can be viewed in an S-Plus session.

4.4.6 Missing Records From Sales Database

Twenty-one records exist in the claims database for which there are no corresponding records for the vehicles' VIN in the sales database. The production year for all of these records has been identified as 1997 in the claims database. This again points to an error: either the sales database is incomplete, or the production date has been incorrectly stated in the claims database. As the number of records is small, this error would have little impact on the subsequent analysis. It has been observed that the vehicles' $VINs$ are within the range assigned to the 1997 records, so it has been decided to include these records. It was felt that making the error of falsely including these claims, and thus slightly increasing the estimated warranty cost is preferable to making the error of falsely excluding them, and underestimating the warranty cost.

Thus, it has been necessary to write additional records to the sales database for these twenty-one records. These have been appended to the sales database, with the $prodn$ field set to -1 to indicate that these records are added. The S-Plus script *AddSale.ssc* has been written to do the processing and is included in Appendix A.5.

4.5 Component Ages in Subsequent Repairs

When a vehicle is repaired, a failed component is replaced by a new component. Thus it is necessary to calculate the age of, and the distance travelled with, this new component. Only on the first repair will age and distance of the component be the same as the age of the vehicle.

In order to calculate the age and distance of a component, the data has been sorted by VIN, then by $part$, and then by $repr$ before being imported. Thus, vehicles are grouped together in subsequent records. Consequently, a record with the same VIN and $part$ as the record immediately preceding it represents a second or subsequent replacement of a component. The age of such components can be calculated by subtracting the repair date of the previous record from that of the current record. The distance travelled with the replaced component can be calculated in a similar fashion, and put in a new field called $dist2$. Although the modelling in the current study does not use $dist2$, it has been included so that it can be used in future studies.

The S-Plus function *fCompAge* has been written to calculate the age of components that were replaced more than once in a vehicle. The script is presented in Appendix A.6. The output of *fCompAge(clean97)* has been put in the data frame *compAge*, which is included on the compact disc, and may be viewed in an S-Plus session.

4.6 Database of Failed and Censored Times

In order to perform survival analysis, a database with the following fields has been created: VIN, $age2$, $dist2$, $part$, $partname$ and $status$, a field to indicate whether the record represented a failed component (1) or a censored observation (0). This database shall be referred to as the survival database. It has been built up from the claims and sales databases, and comprises: (i) components that have failed, (ii) components that have been replaced but had not failed by the end of the observation period, and (iii) vehicles with no warranty claims. The following describes how these three different types of records have been identified.

The records of failed components in the survival database have been obtained from the data frame of component ages (*compAge*). Since each record in *compAge* represents a failed component, all of the records in *compAge* need to be written into the survival database, with *status* set to one.

The second type of record in the survival database is that of replaced components that have not failed by the end of the warranty period. Each entry in *compAge* represents one failed component and one replaced component. A subsequent failure of a replaced component has resulted in another claim being made, which is then represented in the claims database by a another record. The way in which these records are identified has been described in the previous section. The last record for each unique vehicle/part combination represents a component that has not failed. These records would be represented in the survival database as a censored record. The non-failed components' censored ages have been calculated as the difference between the download date, 21/5/2002 and the date of the last repair. The distance travelled by the vehicle since the components were fitted, has been estimated from the usage rate ($dist/age2$) of that vehicle multiplied by the component's age. The fields *part*, *partname*, *status*, the last of which has been set to zero, are included to complete each censored record in the survival database.

The third type of record in the survival database is that of a censored record of a vehicle that has not had a claim. As survival analysis has been performed on a component level, these censored records come from vehicles without a claim on a particular component. For each of these vehicles, a non-failed (censored) record has been written in the survival database. The age of the vehicle has been previously calculated in the sales database (see Section 4.2) and stored in the field $age2$. The field *status* has been set to zero to indicate that these were censored observations. No values have been entered in $dist2$ for these records.

The survival database has then been created by merging records from these three different types. The building up of the survival database as described here has been done by means of an S-Plus function, *fSurvive*, which is presented in Appendix A.7.

4.7 Conclusion

In this chapter, the nature of the data that was available for this study has been described. An initial data exploration has been described, which revealed the need for some data cleaning. The type of errors identified in the data, and how these were treated to maximise the accuracy of the subsequent analysis has also been discussed. To obtain the age of components, the difference between the repair and sale dates has been calculated, and then this value has been checked against the recorded values of age and distance. The details of the processing involved has been discussed in this chapter. The aim of this processing is to build up a database of observations of both failed and non-failed components that is to be used for survival analysis.

The reliability of each component is modelled in the next chapter, as is the cost of its repair, from which a cost model for each component is developed. From these individual models, an overall warranty cost model is procured. The next chapter discusses the theory behind the models, whilst the chapter following that, discusses how the models are used in conjunction with the data described in this chapter to estimate the overall warranty cost.

Chapter 5
Survival and Cost Models

5.1 Introduction

The last chapter described the raw data that was available for this project, and discussed how this data was checked and cleaned. The focus of the current chapter is the theoretical development of warranty cost models. Models for both the reliability of components and the cost of repairing these components are developed, and from these, a model of the warranty cost is constructed. An approach using bootstrap resampling is also discussed. The next chapter discusses the implementation of these models using the manufacturer's data.

This chapter starts by outlining the assumptions made in developing the models. Following that is a discussion of the techniques used in analysing survival or life data, of which warranty data is an example. Then, non-parametric and parametric models of individual component failures are developed. Both point and interval estimates of component reliability are obtained. A discussion of the use of the Renewal Process in warranty cost analysis follows. This is used to obtain point and interval estimates of the expected number of failures of a component during the warranty period. A model of the cost of component repair is then developed. Unlike previous research into warranty cost analysis, the cost of repairing a component is not treated as a constant, but as a random variable. A model of the warranty cost is then developed based on the models of component reliability and cost of repair. Both point and interval estimates are obtained. This approach is then compared to the more direct approach of obtaining a bootstrap point and interval estimate of the total warranty cost. The chapter ends with a discussion of the limitations of the models.

5.2 Assumptions

In this section, a number of assumptions that have been made in producing models of warranty costs are discussed.

5.2.1 Validity of Claims

It has been assumed that all component failures during the warranty period have lead to a claim, and that all claims have been valid. However, this may not have been strictly correct, for there may have been users who had not bothered to make claims for minor faults because it was inconvenient to do so. Conversely, some users may have made claims soon after the warranty period ended, or when only one dimension of the warranty had expired. For example, the distance limitation may have been surpassed, but the time limitation may not have. Dealerships may have been willing to make a warranty claim by adjusting the distance at claim to be within the bounds of the warranty. These variations have been ignored because their numbers are likely to be insignificant compared to the total number of claims.

In addition, the manufacturer's data has been assumed to be genuine and correct, apart from the errors that were described in the last chapter.

There are many instances in the manufacturer's database of claims being honoured after the expiration of the warranty. The manufacturer's representative has indicated that this does occur if a known manufacturing fault exists. However, because owners are free to go to any mechanic after the warranty period, the database does not include all of the post-warranty failure data. Thus, the database is not representative of the total number of failures occurring after the warranty expires, so no data beyond the warranty period has been used in the analysis conducted in this study.

5.2.2 Timing of Failures

In the context of this warranty study, failure of a component has been deemed to occur when a user has made a claim. In reality, this may not strictly be the case, because a component could show signs of impending failure before it actually fails. The customer may then have acted upon these signals rather than have waited until final failure of a component. Christer (1981) and Christer and Waller (1984) developed this concept of a lead-up time into failure, which they called "delay time".

In other instances, a fault may have occurred some time before the claim was made. In the case of minor problems, such as the deterioration of a door bush or the failure of a radio, cassette or compact disc component, the user may have held off making a claim to a more convenient time, such as that of a scheduled service.

When looking at a wide range of components, it is difficult to ascertain the precise timing of the failure of a component. Using these factors makes the modelling more complicated. The timing of the claim, rather than the timing of the failure, is of importance to the manufacturer in warranty analysis because this is when payment of the claim has to be made. Thus, in modelling failure of components in this study, it has been assumed that a failure occurred when a claim was made.

5.2.3 Repair Time

As the vehicle repair time is generally small compared to the mean time to failure of a component, it has been assumed to be zero. This is standard practice in the literature. It has also been assumed that the distance travelled by a vehicle before sale was negligible, and that the number of vehicles that were decommissioned (for example, if the vehicle was written off) in the time interval under investigation, was also negligible.

5.2.4 Nature of Repair

The manufacturer's representative claimed that all repairs under warranty were carried out by replacing faulty parts. Thus, if a part is listed on the claims database, then that particular part had been replaced by a new one. It has been assumed that the part identified in the claims database represents both the part being replaced and the replacement part. It has also been assumed that all repairs were as good as new.

There are a number of records with no part listed. These repairs could include adjustments or cosmetic repairs with no actual part being replaced, and have therefore been analysed separately.

5.2.5 Independence of Component Failures.

In multi-component products, failure of one part may lead to failure of subsequent components. The modelling of such systems is quite complex, as the model must incorporate aspects of the dependent component failures. Blischke and Murthy (1992) advocated the modelling of component failure as independent failures only, for the sake of simplicity, unless there is strong evidence to the contrary, or unless specific dependencies are being investigated. This suggestion has been followed and it has been assumed that the probability of failure of a component is independent of other component failures.

5.2.6 Uniformity of Parts

It has also been assumed that all components bearing a particular part number were identical. This might not have strictly been the case, as manufacturing conditions may vary from batch to batch. It is true to say, however, that a component, as identified by a part number, was supplied by a single supplier with fixed specifications. One could analyse the performance of a component obtained from different suppliers, but that is not the aim of the present investigation. Thus, it has been assumed that all parts with the same part number were identical, and that the time to failure of such components was a random variable derived from the same distribution.

5.2.7 Location Factors

One may build a location factor into a model to cater for the different climatic conditions encountered by users in different locations. Although some inferences can be made about where a vehicle has been driven from both purchase and warranty data, there is no restriction as to where or how the vehicle can actually be driven after purchase. Therefore, a location factor has not been included in the models used in this study, and it has been assumed that data affected by these factors were randomly distributed. Similarly, other vehicle operational factors were not included in the model, and it has been assumed that these factors were randomly distributed in the data.

5.2.8 Dimensionality of Warranty

Most vehicle warranties are based on time and distance, and are therefore, two-dimensional. The warranty in the current study is for three years or 100,000 kilometres, whichever occurs first. Some authors developed models based on two-dimensional distributions, for example, Blischke and Murthy (1992) and (1994), Murthy, Iskandar and Wilson (1995) and Singpurwalla and Wilson (1993). These have been discussed in Section 3.2. Other studies developed models based on usage rates, for example, Chun and Tang (1999), Gertsbakh and Kordonsky (1998) and Moskowitz and Chun (1994), which have also been discussed in Section 3.2. Hu, Lawless and Suzuki (1998) made the assumption that mileage accumulation is linear over time. Lu's (1998) data had an average usage rate of 1000 miles per month, which equates to 19,300 kilometres per year. Lawless, Hu and Cao's (1995) data found that over a large population of customers, the median rate of use was about 13,000 miles, or 20,100 kilometres per year. Our data shows a median usage rate of 19,700 kilometres per year, which lies between these two figures. It would be expected that these two American figures would be similar to our Australian figure, since it is possible to travel vast distances in both countries. With this usage rate, few users would reach 100,000 kilometres by the end of three years. Therefore, the time limitation is the more critical measure of the warranty length. The study data contains no information about the distance travelled by vehicles with no warranty claims, and it was not possible to obtain this information through a survey. Thus, it would have to be estimated, which could introduce a new source of error. Therefore, the warranty modelling in this study is based on time alone. It has been assumed that the number of vehicles that would reach 100,000 kilometres within the three years would not be significant.

Let us examine the consequences of this decision. Using the data for the current investigation, an upper non-outlier limit for usage of 44,700 kilometres per year was obtained. This figure was obtained by using the last claim record of each vehicle in the claims database, and eliminating records containing outliers in vehicle's ages ($age2$) or distance ($dist$) travelled. In addition, records of vehicles less than two months old, or of vehicles that had travelled less than 3,300 kilometres (the average distance travelled by the vehicle in two months), were also eliminated before calculating this non-outlier limit for usage. The usage figure of 44,700 kilometres per year is a reasonable figure for a heavy user such as a sales representative. Using this figure, 11% of the users that made a claim reached the distance limitation of the warranty (100,000 kilometres) before they reached the time limitation. (These figures were calculated in the S-Plus function *fUsage*, which is presented in Appendix B.1.) Figure 5.1 shows vehicle usage rates of cars with warranty claims. Although the data shown in the figure has been taken from a limited sample, that is the claims data only, it does support the notion that the vast majority of vehicles will reach their time limitation of the warranty before they reach their distance limitation. Ideally, Figure 5.1 should include vehicles without claims. However, the resources to conduct a survey to obtain the data were not available in this study.

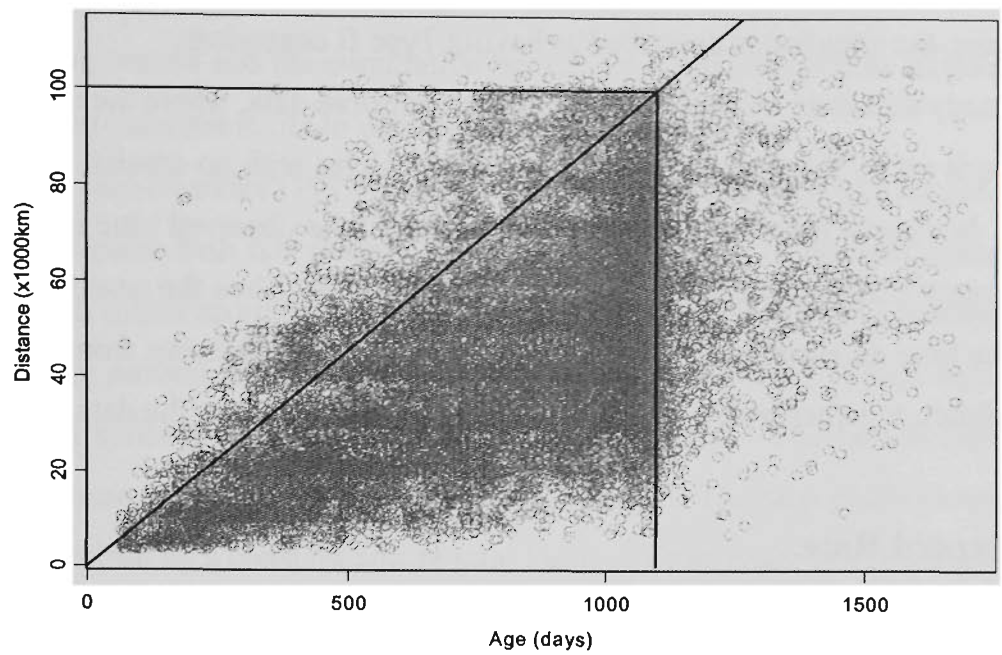


Fig. 5.1. *Vehicle usage rates, showing warranty limits and max usage rate for the full 3-year warranty (diagonal line).*

5.3 Survival Analysis Methods

Warranty data are an example of life data. Thus, the techniques used to analyse life data are used to analyse warranty data. The concepts of censoring times, survival times, and hazard rate, or instantaneous rate of failure, which are used in survival analysis are discussed in the sections that follow. The concepts discussed below can be found in many reliability texts, such as Lawless (1982) and Nelson (1982).

5.3.1 Censoring Times

Survival analysis concerns the lifetime distribution of items varying from mechanical or electronic components to humans. In this chapter, survival analysis techniques are used to estimate the failure distribution, both non-parametric and parametric, of components. It is not practical to observe the lifetime of items until all items fail or die. Usually, the time of a study is limited to a set period of observation. This results in the censoring of lifetimes to the observation limits. There are two types of censoring. When the lifetime of an item is observed for a fixed period of time, t_c, the situation is described as having Type I censoring. Within this type of censoring, if the failure time, T, is known to be greater than the observation time, t_c, so that $T \geqslant t_c$, the censoring is said to be *right-censored.* When the lifetime of an item is observed until a set number, r, of events (usually death or failure) occur, the situation is described as having Type II censoring.

Warranty data is an example of Type I right-censored data, where the upper limit of observation is set by the warranty period, t_w. For vehicles with no repairs, the observed time is t_w. In cases where vehicles have been repaired, the observed time is the balance of the warranty period, after the last repair on the vehicle. Thus the resulting database includes the time an event occurs or the time the item was withdrawn from observation (censored time), together with a status indicating to which category the data belongs.

5.3.2 Hazard Rate

The hazard function is defined as the instantaneous rate of failure.

Let T be the random variable representing the lifetime of interest. Then T will usually take some finite value, t, in the interval $[0, \infty)$. The cumulative failure distribution is given by $F(t)$ and the density function is given by $f(t)$. The survival, or reliability function, which is the probability of surviving to a time $T > t$ is given by $S(t) = 1 - F(t) = \Pr(T > t)$. The hazard function is defined as

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0} \frac{\Pr(t \leqslant T < t + \Delta t | T \geqslant t)}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{\Pr(t \leqslant T < t + \Delta t)}{\Delta t} \div \Pr(T \geqslant t) \\ &= \frac{f(t)}{S(t)}. \end{aligned}$$

Thus, $h(t)\delta t$ is interpreted as the probability that the item will fail in $[t, t + \delta t)$, given that it has not failed before t.

The cumulative hazard function is given by

$$\begin{aligned} H(t) &= \int_0^t h(x)\, dx \\ &= \int_0^t \frac{f(x)}{S(x)} dx \\ &= \int_0^t \frac{f(x)\, dx}{1 - F(x)} \\ &= \int_0^t \frac{dF(x)}{1 - F(x)} \\ &= \int_0^t \frac{-dS(x)}{S(x)}, \end{aligned}$$

giving

$$H(t) = -\log S(t)\,. \tag{5.1}$$

Both the hazard and the cumulative hazard functions are used in establishing non-parametric and parametric fits to data in the following sections.

Murthy and Blischke (1992b) stated that many products exhibit a hazard rate which has a characteristic bath-tub shaped curve (see Figure 5.2). In the first phase (0 to t_1), there is a high infant mortality rate. Young items fail due to defective materials or poor manufacturing processes (often referred to as teething problems). These failures can be reduced by a suitable burn-in period and testing program. The failure rate decreases during this phase as these initial failures are weeded out. The next phase (between t_1 and t_2) is characterised by a constant failure rate. Failure in this phase is due purely to chance, and not age. This is characteristic of electrical and electronic components. During the last phase (beyond t_2), the failure rate increases as the item ages. This is characteristic of mechanical components that are wearing out.

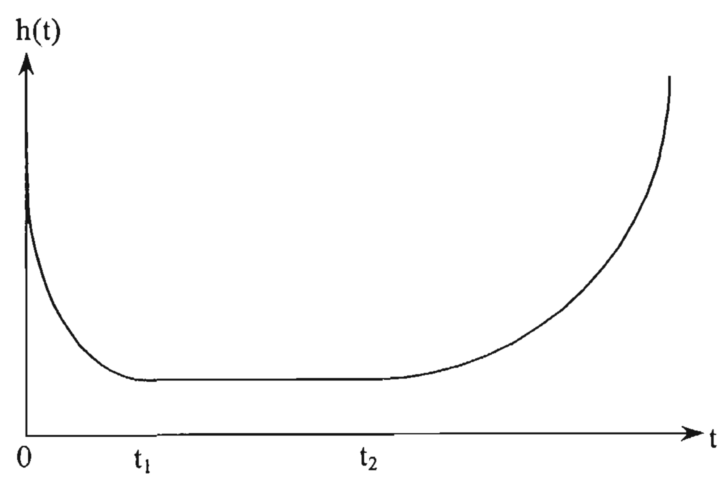


Fig. 5.2. *Bath tub hazard rate. (Adapted from Fig. 1, Murthy and Blischke, 1992b.)*

Ascher and Feingold (1984) distinguished between the bathtub curve for the hazard rate (which they call force of mortality) of parts or simple products, and the rate of occurrence of failures of a single repairable system. Figure 5.2 and the above discussion refer to the former. The modelling of multi-component items can be done by modelling each component failure and then relating each component failure to item failure. This would depend upon how a component's failure affects the failure of other components and of the item itself. Sometimes the failure of a component can damage or weaken other components. Thus, the modelling of failure interactions can be very difficult. Because of the large number of components involved in this study, component failures have been taken to be independent of one another.

5.4 Non-Parametric Estimation of Component Reliability

The use of non-parametric models for survival data is a newer approach than its parametric counterpart. The non-parametric approach is appropriate when a parametric form cannot be found, or is not required. It can be used to interpolate results within the bounds of the data.

The number of items at risk, $k(t)$, is the number of items in the sample that have not failed just before time t. Let the period of observation, which in our case is the warranty length, t_n, be divided into intervals so that $0 = t_0 < t_1 < t_2 < \cdots < t_n$. Then interval i is $[t_{i-1}, t_i)$, where $1 \leqslant i \leqslant n$. An estimate of the probability of an item surviving interval i is

$$p_i = \frac{k(t_i) - d_i}{k(t_i)},$$

where d_i is the number of failures, or deaths, during interval i. Using the chain rule for conditional probability over all the intervals, the survival to t_n is

$$\begin{aligned} S(t_n) &= S(t_n|t_{n-1}) \times S(t_{n-1}|t_{n-2}) \times \ldots S(t_1|t_0) \times S(t_0) \\ &= p_n \times p_{n-1} \times \ldots p_1 \times p_0 \\ &\approx \prod_{i=1}^{n} p_i, \end{aligned}$$

where $p_0 = S(t_0) = 1$. This leads to the Kaplan-Meier estimator of survival,

$$\hat{S}_K(t) = \prod_{t_i \leqslant t} \frac{k(t_i) - d_i}{k(t_i)}.$$

It is also known as the product-limit estimator. There are numerous references on the Kaplan-Meier estimator, such as Venables and Ripley (1999).

The Actuarial Method, as detailed in Nelson (1982), is a similar approach to the above, but uses an adjustment term to estimate survival at the centre of an interval. It assumes that items are taken out of observation at the centre of an interval. The survival of an item in the interval i is

$$p_i = 1 - \frac{d_i}{k(t_i) - 0.5r_i},$$

where r_i is the number of items removed from observation during interval i. This leads to an estimator of survival

$$\hat{S}_N(t) = \prod_{t_i \leqslant t} \left(1 - \frac{d_i}{k(t_i) - 0.5r_i}\right).$$

The Actuarial Method is used to construct life tables of human life expectancy.

A different approach was taken by Fleming and Harrington (1984). An estimate of the cumulative hazard is given by

$$\hat{H}(t) = \sum_{t_i \leqslant t} \frac{d_i}{k(t_i)}.$$

This is known as the Nelson estimator of the cumulative hazard. Using (5.1), we have $S(t) = \exp(-H(t))$. This leads to an estimate of survival as

$$\hat{S}_F(t) = \exp\left(-\sum_{t_i \leqslant t} \frac{d_i}{k(t_i)}\right).$$

With large data samples, the Fleming-Harrington estimator produces similar results to the Kaplan-Meier estimator.

An approach given in Jardine (1984) is a little different to the above. It uses the average number of items at risk in an interval by averaging the number at risk at the start and end of an interval, that is, $v(t_i) = \frac{1}{2}(k(t_i) + k(t_{i+1}))$. The average hazard rate for the interval is then calculated as $d_i/v(t_i)$. The cumulative hazard to the end of an interval is then calculated as the sum of these hazards. Equation (5.1) leads to $S(t) = \exp\{-H(t)\}$, from which we are able to estimate the survival of each interval. The estimator is

$$\hat{S}_J(t) = \exp\left(-\sum_{t_i \leqslant t} \frac{d_i}{v(t_i)}\right).$$

Each of the above methods can produce slightly different results. In a warranty application, the number of failures is small compared to the total number produced, so the above methods would yield very similar results. As the Kaplan-Meier method is the most commonly used method, and since it is readily implemented in S-Plus, it has been used in obtaining a non-parametric fit to the data.

Having established a method of obtaining a point estimate for survival, let us consider the variance of the estimate. The literature contains a number of suggestions and modifications for the variance of a survival estimate, and these are reviewed in Venables and Ripley (1999). Greenwood's formula for the variance of the survival function estimate is obtained as follows. Since $\hat{H}(t)$ is the sum of independent increments, the variance of

$\hat{H}(t)$ is the cumulative sum of terms. That is,

$$var\left(\hat{H}(t)\right) = \sum_{t_i \leqslant t} \frac{d_i}{k(t_i).k(t_i - d_i)}.$$

This leads to an expression for the variance of $\hat{S}(t)$:

$$var\left(\hat{S}(t)\right) = \hat{S}(t)^2 \sum \frac{d_i}{k(t_i).k(t_i - d_i)}.$$

This is again in common usage and is readily implemented in S-Plus, so this has been used for the variance of the survival function.

5.5 Parametric Models of Component Reliability

Parametric models for survival data are becoming less popular now that non-parametric approaches are available. This is because no assumptions need to be made about the parametric form using the non-parametric approach. However, parametric modelling is the only appropriate approach when extrapolation beyond the data is necessary. This is the case when future trends need to be predicted. In warranty analysis, if the cost of extending a warranty is to be estimated, as is the case in this study, a parametric model is required.

The simplest parametric model is the exponential model, which has a constant hazard rate. It has often been used in reliability analysis because of its mathematical tractability. As Hill and Blischke (1987) put it: "closed-form results are most easily obtained if it is assumed that the lifetime of the items in question are exponential." It is an appropriate model where few failures have been observed and when initial burn-in occurs. Since the warranty period is short compared to the lifetime of a component, we are observing the lifetime of a component during the flat part of the bath tub hazard curve, again making the exponential model appropriate.

However, Blischke and Murthy (2000) recognised the flexibility of the Weibull distribution and have stated that it is "the most widely used failure distribution in reliability applications" (page 108). Although the Weibull does offer greater flexibility than the exponential distribution, both models are used in this study. The exponential's tractability, especially when obtaining confidence limits, is the reason it has been used.

The rest of this section discusses the use of the likelihood function to obtain parameter estimates of models. Point and variance estimates of the exponential model are discussed, followed by a similar discussion of the Weibull model, both of which are used to model the reliability of components. Following this, the next section examines how the point estimates and their variances can be used to obtain an estimate of the expected number of claims made during the warranty period.

5.5.1 Log Likelihood Function

The parameters of a distribution used to model data can be obtained using maximum likelihood methods. This approach is used in many texts, for example, Cox and Oakes (1984). For a large sample size, the maximum likelihood estimators have good statistical properties, and are readily calculated. The cumulative distribution function of a maximum likelihood estimator is close to a normal distribution with a mean equal to the parameter being estimated, and with a variance no greater than that of other estimators.

Let $T_1, T_2, \ldots T_n$ be a random sample of size n of ages to failure, drawn from a probability density function $f(t_j;\theta)$, where t_j is the time of the jth failure and θ is an unknown parameter set. The likelihood function of this random sample as a function of the parameter is given by

$$L(t_j;\theta) = \left(\prod_{j\in F} f(t_j;\theta)\right)\left(\prod_{j\in C} S(t_j;\theta)\right), \tag{5.2}$$

where $f(t_j)$ is the density distribution of T, $S(T)$ is the survival function, F is the set of observations reaching failure and C is the set of censored observations. The parameters of the model can be found by maximising L. To simplify the calculations, the logarithm of the likelihood function is first taken, as the maximum of L can be obtained by maximising $\log(L)$. From (5.2), the log likelihood function is

$$\begin{aligned} l(t_j;\theta) &= \log(L) \\ &= \log\left(\prod_{j\in F} f(t_j;\theta)\right) + \log\left(\prod_{j\in C} S(t_j;\theta)\right), \end{aligned}$$

which can be written as

$$l(t_j;\theta) = \sum_{j\in F}\log(f(t_j;\theta)) + \sum_{j\in C}\log(S(t_j;\theta)). \tag{5.3}$$

5.5.2 Exponential Model

This subsection discusses both point and interval estimates of the parameter of an exponential model fitted to a set of data. The next subsection does the same for a Weibull model.

Point Estimate

The probability density function for the exponential distribution is

$$f(t) = \lambda e^{-\lambda t}, \tag{5.4}$$

where the parameter λ is the failure rate. The cumulative distribution function and the survival function are given by

$$F(t) = 1 - e^{-\lambda t}, \text{ and}$$

$$S(t) = 1 - F(t) = e^{-\lambda t}, \tag{5.5}$$

respectively. Using (5.4) and (5.5), substitute for $f(t_j;\theta)$ and $S(t_j;\theta)$ in (5.3):

$$\begin{aligned} l(t_j;\theta) &= \sum_{j\in F} \log\left(\lambda e^{-\lambda t_j}\right) + \sum_{j\in C} \log\left(e^{-\lambda t_j}\right) \\ &= \sum_{j\in F} (\log\lambda - \lambda t_j) - \sum_{j\in C} \lambda t_j \\ &= \sum_{j\in F} \log\lambda - \sum_{j\in F\cup C} \lambda t_j. \end{aligned}$$

If r of the sample of n items reach failure, so that $n - r$ of the sample are censored items, we have

$$\begin{aligned} l(t_j;\theta) &= \sum_{j=1}^{r} \log\lambda - \sum_{j=1}^{n} \lambda t_j \\ &= r\log\lambda - \lambda\sum_{j=1}^{n} t_j. \end{aligned}$$

A local maximum of l can be found when $\partial l(t_j;\theta)/\partial\lambda = 0$. Thus, the maximum likelihood estimator of the parameter λ can be obtained by solving

$$\frac{r}{\hat{\lambda}} - \sum_{j=1}^{n} t_j = 0,$$

which yields an expression for the maximum likelihood estimate of λ as

$$\hat{\lambda} = \frac{r}{\sum_{j=1}^{n} t_j}. \tag{5.6}$$

Variance

Nelson (1982, p318) gives the maximum likelihood estimate of the variance of the parameter θ ($= 1/\lambda$) of the exponential distribution (5.4) for time-censored data as

$$V\hat{a}r\left(\hat{\theta}\right) = \frac{\hat{\theta}^2}{\sum_j \left[1 - \exp\left(-t_j/\hat{\theta}\right)\right]},$$

where t_j is the censoring time for item j. From this expression we can obtain the variance of the failure rate as

$$V\hat{a}r\left(\hat{\lambda}\right) = \frac{\hat{\lambda}^2}{\sum_j \left[1 - \exp\left(-t_j\hat{\lambda}\right)\right]}.$$

5.5.3 Weibull Model

This section discusses point and interval estimates of the parameters of a Weibull model fitted to a set of data.

Point Estimate

The Weibull distribution may be written in the form

$$f(t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} \exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right], \tag{5.7}$$

where β is the shape parameter and η is the scale parameter. It follows that the survival function is

$$S(t) = \exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right]. \tag{5.8}$$

Substituting for $f(t_j;\theta)$ and $S(t_j;\theta)$ in (5.3), we get

$$\begin{aligned} l(t_j;\theta) &= \sum_{j\in F} \log\left[\frac{\beta}{\eta}\left(\frac{t_j}{\eta}\right)^{\beta-1} \exp\left[-\left(\frac{t_j}{\eta}\right)^{\beta}\right]\right] + \sum_{j\in C} \log\left[\exp\left(-\left(\frac{t_j}{\eta}\right)^{\beta}\right)\right] \\ &= \sum_{j\in F}\left[\log\frac{\beta}{\eta} + (\beta-1)\log\frac{t_j}{\eta} - \left(\frac{t_j}{\eta}\right)^{\beta}\right] - \sum_{j\in C}\left(\frac{t_j}{\eta}\right)^{\beta} \\ &= \sum_{j\in F}\log\frac{\beta}{\eta} + (\beta-1)\sum_{j\in F}(\log t_j - \log\eta) - \sum_{j\in F\cup C}\left(\frac{t_j}{\eta}\right)^{\beta}. \end{aligned}$$

If r of the sample of n components reach failure, so that $n-r$ observations are censored items, we have

$$\begin{aligned} l\left(t_j;\theta\right) &= r\log\frac{\beta}{\eta}+(\beta-1)\sum_{j=1}^{r}\left(\log t_j-\log\eta\right)-\sum_{j=1}^{n}\left(\frac{t_j}{\eta}\right)^{\beta} \\ &= r\log\beta-r\log\eta+(\beta-1)\sum_{j=1}^{r}\log t_j-(\beta-1)\,r\log\eta-\eta^{-\beta}\sum_{j=1}^{n}t_j^{\beta} \\ &= r\log\beta-r\beta\log\eta+(\beta-1)\sum_{j=1}^{r}\log t_j-\eta^{-\beta}\sum_{j=1}^{n}t_j^{\beta}. \end{aligned}$$

A local maximum of l is obtained by taking partial derivatives with respect to each parameter, β and η, and then setting them to zero. Now

$$\begin{aligned} \frac{\partial\, l\left(t_j;\theta\right)}{\partial\,\beta} &= \frac{r}{\beta}-r\log\eta+\sum_{j=1}^{r}\log t_j+\eta^{-\beta}\log\eta\sum_{j=1}^{n}t_j^{\beta}-\eta^{-\beta}\sum_{j=1}^{n}t_j^{\beta}\log t_j \\ &= \frac{r}{\beta}-r\log\eta+\sum_{j=1}^{r}\log t_j-\eta^{-\beta}\sum_{j=1}^{n}t_j^{\beta}\log\frac{t_j}{\eta}, \end{aligned} \tag{5.9}$$

and

$$\begin{aligned} \frac{\partial\, l\left(t_j;\theta\right)}{\partial\,\eta} &= -\frac{r\beta}{\eta}+\beta\eta^{-\beta-1}\sum_{j=1}^{n}t_j^{\beta} \\ &= \frac{\beta}{\eta}\left(-r+\eta^{-\beta}\sum_{j=1}^{n}t_j^{\beta}\right). \end{aligned}$$

Equating $\dfrac{\partial\, l\left(t_j;\theta\right)}{\partial\,\eta}$ to zero, we obtain the relationship between the parameters:

$$\hat{\eta}^{\hat{\beta}}=\frac{1}{r}\sum_{j=1}^{n}t_j^{\hat{\beta}},$$

yielding

$$\hat{\eta}=\left(\frac{1}{r}\sum_{j=1}^{n}t_j^{\hat{\beta}}\right)^{\frac{1}{\hat{\beta}}}. \tag{5.10}$$

Equating $\partial l\left(t_j;\theta\right)/\partial\beta$ to zero and substituting for $\hat{\eta}$ in (5.9), we have

$$\frac{r}{\hat{\beta}}-r\log\left(\frac{1}{r}\sum_{j=1}^{n}t_j^{\hat{\beta}}\right)^{\frac{1}{\hat{\beta}}}+\sum_{j=1}^{r}\log t_j-\left(\left(\frac{1}{r}\sum_{j=1}^{n}t_j^{\hat{\beta}}\right)^{\frac{1}{\hat{\beta}}}\right)^{-\hat{\beta}}\sum_{j=1}^{n}t_j^{\hat{\beta}}\log\frac{t_j}{\left(\frac{1}{r}\sum_{j=1}^{n}t_j^{\hat{\beta}}\right)^{\frac{1}{\hat{\beta}}}}=0$$

$$\frac{r}{\hat{\beta}}-r\log\left(\frac{1}{r}\sum_{j=1}^{n}t_j^{\hat{\beta}}\right)^{\frac{1}{\hat{\beta}}}+\sum_{j=1}^{r}\log t_j-\left(\frac{1}{r}\sum_{j=1}^{n}t_j^{\hat{\beta}}\right)^{-1}\sum_{j=1}^{n}t_j^{\hat{\beta}}\left(\log t_j-\log\left(\frac{1}{r}\sum_{j=1}^{n}t_j^{\hat{\beta}}\right)^{\frac{1}{\hat{\beta}}}\right)=0$$

$$\frac{r}{\hat{\beta}}-\frac{r}{\hat{\beta}}\log\left(\frac{1}{r}\sum_{j=1}^{n}t_j^{\hat{\beta}}\right)+\sum_{j=1}^{r}\log t_j-r\left(\sum_{j=1}^{n}t_j^{\hat{\beta}}\right)^{-1}\sum_{j=1}^{n}t_j^{\hat{\beta}}\left(\log t_j-\frac{1}{\hat{\beta}}\log\left(\frac{1}{r}\sum_{j=1}^{n}t_j^{\hat{\beta}}\right)\right)=0$$

$$\frac{r}{\hat{\beta}}-\frac{r}{\hat{\beta}}\log\left(\frac{1}{r}\sum_{j=1}^{n}t_j^{\hat{\beta}}\right)+\sum_{j=1}^{r}\log t_j-r\left(\sum_{j=1}^{n}t_j^{\hat{\beta}}\right)^{-1}\sum_{j=1}^{n}t_j^{\hat{\beta}}\log t_j$$
$$+\frac{r}{\hat{\beta}}\left(\sum_{j=1}^{n}t_j^{\hat{\beta}}\right)^{-1}\sum_{j=1}^{n}t_j^{\hat{\beta}}\log\left(\frac{1}{r}\sum_{j=1}^{n}t_j^{\hat{\beta}}\right)=0$$

$$\frac{r}{\hat{\beta}}-\frac{r}{\hat{\beta}}\log\left(\frac{1}{r}\sum_{j=1}^{n}t_j^{\hat{\beta}}\right)+\sum_{j=1}^{r}\log t_j-r\left(\sum_{j=1}^{n}t_j^{\hat{\beta}}\right)^{-1}\sum_{j=1}^{n}t_j^{\hat{\beta}}\log t_j+\frac{r}{\hat{\beta}}\log\left(\frac{1}{r}\sum_{j=1}^{n}t_j^{\hat{\beta}}\right)=0,$$

which leads to

$$\frac{r}{\hat{\beta}}+\sum_{j=1}^{r}\log t_j-r\left(\sum_{j=1}^{n}t_j^{\hat{\beta}}\right)^{-1}\sum_{j=1}^{n}t_j^{\hat{\beta}}\log t_j=0. \tag{5.11}$$

A value for $\hat{\beta}$ can be found by solving (5.11) iteratively. Once a value for $\hat{\beta}$ is determined, a value for $\hat{\eta}$ can be found using (5.10). Thus, parameter estimates for both parameters of the Weibull distribution can be found using the method of maximum log likelihood.

Variance

For items with a failure density $f\left(x_,;\mu,\sigma\right)$, Nelson (1982, pages 375 to 379) gives the Fisher information matrix for the log likelihood as

$$\mathbf{F}=\begin{bmatrix}-\partial^2 l/\partial\mu^2 & -\partial^2 l/\partial\mu\partial\sigma\\ -\partial^2 l/\partial\mu\partial\sigma & -\partial^2 l/\partial\sigma^2\end{bmatrix}.$$

He gives the large-sample asymptotic covariance matrix of $\hat{\mu}$ and $\hat{\sigma}$ as the inverse of this, namely,

$$var\left(\hat{\theta}\right)=\begin{bmatrix}var\left(\hat{\mu}\right) & cov\left(\hat{\mu},\hat{\sigma}\right)\\ cov\left(\hat{\mu},\hat{\sigma}\right) & var\left(\hat{\sigma}\right)\end{bmatrix}.$$

Thus, for the Weibull distribution, the covariance matrix from a maximum likelihood fit of a set of data is given by

$$var\left(\hat{\theta}\right) = \begin{bmatrix} var\left(\hat{\eta}\right) & cov\left(\hat{\eta},\hat{\beta}\right) \\ cov\left(\hat{\eta},\hat{\beta}\right) & var\left(\hat{\beta}\right) \end{bmatrix}. \tag{5.12}$$

A confidence region for the combination of these parameters can be obtained. As Draper and Smith (1966, page 95) pointed out, it is incorrect to use a rectangular confidence region that is made up of the confidence intervals of the two separate parameters.

Cook and Weisberg (1994, pp218-9) have shown that for a two-parameter model,

$$y|x = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \varepsilon,$$

with independent normal errors having mean 0 and constant variances σ^2, a joint $100(1-\alpha)\%$ confidence region for (θ_1, θ_2) can be obtained from a sample of size n. It is the set of all values of the 2×1 vector θ that satisfies the inequality

$$\left(\theta - \hat{\theta}\right)^T \left[var\left(\hat{\theta}\right)\right]^{-1} \left(\theta - \hat{\theta}\right) \leqslant 2F\left(1-\alpha, 2, n-3\right), \tag{5.13}$$

where $F\left(1-\alpha, 2, n-3\right)$ is the percentage point of the F-distribution with 2 and $n-3$ degrees of freedom that leaves an area of α under the right tail, and $v\hat{a}r\left(\hat{\theta}\right)$ is an estimate of the variance-covariance matrix. The points that satisfy this inequality fall inside this region with centre at $\hat{\theta}$. Thus, using (5.13), a confidence region for the Weibull parameter set can be obtained from the point estimates of the parameters and their variance-covariance matrix. This can then be used to obtain point and interval estimates of the expected number of claims during the warranty period, as discussed in the next section.

5.6 Modelling Failure Using a Renewal Process

This section discusses how the renewal equation can be used to model the number of replacements of a component in a vehicle. Firstly, the formulation of the problem using the renewal function is presented, followed by the solution of the renewal equation for the exponential model. The section ends with a presentation of numerical methods to solve the renewal equation for the Weibull model.

5.6.1 Formulation

When a component fails, it is replaced by a new one. However, the warranty does not start anew on the replaced component. The duration of the warranty is based on the age of the vehicle, not that of the component. The process of replacing a failed component with a new one is called a renewal process. The formulation of the renewal process has been documented in a number of references, such as Cox (1982) and Tijms (1994).

During a warranty, the original item may fail and be replaced with a new one. This process is repeated as many times as required during the warranty period. The lifetime of the original item, and each of its replacements if failure occurs, is a random variable. Let the age at failure of item j be X_j. The X_j are assumed to be independent and identically distributed, with a probability distribution $F(x)$. Let the number of renewals in the time interval $(0, t)$ be $N(t)$, with $N(0) = 0$. This means that the original component is not taken as a renewal.

Now X_j represents the time between the $(j-1)$th and the jth renewal. Then $N(t) \geqslant n$ if and only if $S_n < t$, where $S_n = \sum_{j=1}^{n} X_j$. Since S_n is the sum of n independent and identically distributed random variables, the distribution of S_n is the n-fold convolution of $F(x)$ with itself. Thus the probability that there are n renewals up to time t is given by

$$\Pr(N(t) = n) = \Pr(N(t) \geqslant n) - \Pr(N(t) \geqslant n+1), \tag{5.14}$$

so

$$\Pr(N(t) = n) = \Pr(S_n < t) - \Pr(S_{n+1} < t).$$

Let $M(t)$ be the expected number of replacements of failed components to time t, that is, $M(t) = E[N(t)]$. Given that the first replacement occurs at $x_1 = x$, we have

$$M(t|x_1 = x) = \begin{cases} 0 & \text{if } x \geqslant t \\ M(t-x) + 1 & \text{if } x < t. \end{cases}$$

Thus we have

$$\begin{aligned} M(t) &= \int_0^\infty M(t|x = x)\, dF(x) \\ &= \int_0^t \{1 + M(t-x)\}\, dF(x), \end{aligned}$$

which gives

$$M(t) = F(t) + \int_0^t M(t-x)\, dF(x). \tag{5.15}$$

By using the change of variable rule of integration, (5.15) can be written as

$$\int_0^t M(t-x)\,dF(x) = M(t-x)\,F(x)|_{x=0}^t - \int_0^t F(x)\,dM(t-x). \tag{5.16}$$

Since $F(0)=0$ for a lifetime distribution, and $M(0)=0$, we have from (5.16)

$$\int_0^t M(t-x)\,dF(x) = -\int_0^t F(x)\,dM(t-x).$$

Further, putting $s=t-x$ we have $dM(t-x) = -dM(s)$, so that

$$\begin{aligned}\int_0^t M(t-x)\,dF(x) &= -\int_{s=t}^0 F(t-s)\,dM(s)\\ &= \int_0^t F(t-x)\,dM(x).\end{aligned}$$

Thus, (5.15) can be written as

$$M(t) = F(t) + \int_0^t F(t-x)\,dM(x). \tag{5.17}$$

Let the distribution of $S_n = \sum_{j=1}^n X_j$ be $F_n(t)$, so that $F_n(t) = \Pr(N(t) \geqslant n)$. Since $M(t)$ is the expected number of renewals to time t, we have

$$\begin{aligned}M(t) &= E[N(t)]\\ &= \sum_{n=1}^{\infty} n \Pr[N(t)=n].\end{aligned}$$

Using (5.14), we have

$$\begin{aligned}M(t) &= F_1(t) - F_2(t) + 2F_2(t) - 2F_3(t) + 3F_3(t) - 3F_4(t) + \ldots\\ &= \sum_{n=1}^{\infty} F_n(t).\end{aligned}$$

Now $F_n(t)$ is the n-fold convolution of $F(t)$ with itself, so that

$$\begin{aligned}F_1(t) &= \int_0^t dF(s) = F(t)\\ F_n(t) &= \int_0^t F_{n-1}(t-s)\,dF(s), \quad n=2,3,\ldots\end{aligned}$$

For only a few distributions (for example, exponential and gamma) can (5.15) be formulated algebraically. For most distributions, it has to be solved numerically. Equation (5.15) is firstly solved when $F(t)$ is exponential, and then its numerical solution for the Weibull distribution is discussed in the subsequent subsection.

5.6.2 Solution for the Exponential Distribution

Let us first consider the Laplace transform of $f(t)$, which is defined as

$$L\{f(t)\} = f^*(s) = \int_0^\infty e^{-st} f(t)\, dt.$$

For the exponential distribution, we have $f(t) = \lambda e^{-\lambda t}$, so that after some simple algebra we get

$$f^*(s) = \frac{\lambda}{s+\lambda}. \tag{5.18}$$

Taking the derivative of (5.15) and using Leibnitz's rule, we have

$$\begin{aligned} \frac{dM(t)}{dt} &= \frac{dF(t)}{dt} + \int_0^t \frac{d[M(t-x)\, dF(x)]}{dt} + M(0)\, dF(t) \\ m(t) &= f(t) + \int_0^t m(t-x) f(x)\, dx, \end{aligned}$$

since $M(0) = 0$.

Taking the Laplace transform we get upon using the convolution theorem (Cox, 1982)

$$m^*(s) = f^*(s) + f^*(s)\, m^*(s)$$

and so

$$m^*(s) = \frac{f^*(s)}{1 - f^*(s)}. \tag{5.19}$$

Using (5.18) in (5.19), for the exponential distribution $m^*(s)$ is

$$\begin{aligned} m^*(s) &= \frac{\lambda}{(s+\lambda)\left(1 - \frac{\lambda}{s+\lambda}\right)} \\ &= \frac{\lambda}{s}. \end{aligned}$$

Taking the inverse Laplace transform of this gives $m(t) = \lambda$, and so since $M(t) = \int_o^t m(t)\, dt$,

$$M(t) = \lambda t, \tag{5.20}$$

which, of course, is a familiar result. Solution (5.20) may also be obtained from (5.17) by substituting $F(t) = e^{\lambda t}$, rearranging, differentiating and solving the resulting first order differential equation. This technique is, however, not applicable to more general distributions and will not be dealt with further.

From (5.20), an estimate of the variance of the number of renewals is given by

$$\hat{var}[M(t)] = t^2 \times \hat{var}(\lambda). \tag{5.21}$$

As discussed in Section 5.5.2, an estimate of the variance of λ can be obtained from the deviations, which can then be used to obtain an estimate of the variance of the expected number of renewals, using (5.21).

5.6.3 Numerical Solution for the Weibull Distribution

An approach developed in Xie (1989) to obtain a numerical solution to the renewal equation is used in this study, and is shown below for completeness.

Equation (5.15) can be written as $M(t) = F(t) + \int_0^t M(t-x) f(x)\,dx$, where $f(x)$ is the density of $F(x)$. This equation is known as the Volterra integral equation of the second kind. Xie (1989) stated that many studies have presented numerical methods for solving this equation (for example, Baker, 1977), however, he pointed out that solving (5.17) is simpler.

By partitioning an interval $[a,b]$ into n intervals of equal width, we can use the midpoint rule to evaluate the Stieltjes integral

$$\int_a^b f(x)\,dg(x) \approx \sum_{i=1}^{n} f\left(x_{i-1/2}\right)\left\{g(x_i) - g(x_{i-1})\right\}, \tag{5.22}$$

where $x_{i-1/2} = \frac{1}{2}(x_i + x_{i-1})$.

For a given t, let the partition of the interval $[0,t]$ for $0 \leqslant i \leqslant n$ be $t_0 < t_1 < t_2 < \cdots < t_n$, with $t_0 = 0$ and $t_n = t$.

Then using (5.17), we have

$$M(t_j) = F(t_j) + \int_0^{t_j} F(t_j - x)\,dM(x).$$

Using (5.22), $M(t_j)$ can be approximated by

$$\begin{aligned} M(t_j) &\approx F(t_j) + \sum_{i=1}^{j} F\left(t_j - t_{i-\frac{1}{2}}\right)\left\{M(t_i) - M(t_{i-1})\right\} \\ &= F(t_j) + \sum_{i=1}^{j-1} F\left(t_j - t_{i-\frac{1}{2}}\right)\left\{M(t_i) - M(t_{i-1})\right\} + F\left(t_j - t_{j-\frac{1}{2}}\right)\left\{M(t_j) - M(t_{j-1})\right\} \end{aligned}$$

Thus,

$$\begin{aligned} \left\{1 - F\left(t_j - t_{j-\frac{1}{2}}\right)\right\} M(t_j) \approx F(t_j) &+ \sum_{i=1}^{j-1} F\left(t_j - t_{i-\frac{1}{2}}\right)\left\{M(t_i) - M(t_{i-1})\right\} \\ &- F\left(t_j - t_{j-\frac{1}{2}}\right) M(t_{j-1}), \end{aligned}$$

giving

$$M(t_j) \approx \frac{F(t_j) + \sum_{i=1}^{j-1} F\left(t_j - t_{i-\frac{1}{2}}\right)\{M(t_i) - M(t_{i-1})\} - F\left(t_j - t_{j-\frac{1}{2}}\right) M(t_{j-1})}{1 - F\left(t_j - t_{j-\frac{1}{2}}\right)}$$

and so

$$M(t_j) \approx \frac{F(t_j) + S_j - F\left(t_j - t_{j-\frac{1}{2}}\right) M(t_{j-1})}{1 - F\left(t_j - t_{j-\frac{1}{2}}\right)}, \tag{5.23}$$

where

$$S_j = \sum_{i=1}^{j-1} F\left(t_j - t_{i-\frac{1}{2}}\right)\{M(t_i) - M(t_{i-1})\}. \tag{5.24}$$

Thus, $M(t_j)$ can be calculated recursively. Similarly, S_j can also be calculated recursively using $M(t_{j-1})$ and the value of $F(t)$ at the midpoint of the interval. Now

$$\begin{aligned} t_j - t_{j-\frac{1}{2}} &= t_j - \frac{1}{2}(t_j + t_{j-1}) \\ &= \frac{1}{2}(t_j - t_{j-1}) \\ &= \frac{1}{2}h, \end{aligned}$$

where h is the interval width. Thus, (5.23) can be simplified to

$$M(t_j) \approx \frac{F(t_j) + S_j - F\left(\frac{1}{2}h\right) M(t_{j-1})}{1 - F\left(\frac{1}{2}h\right)}. \tag{5.25}$$

Xie (1989) provided an implementation of (5.25) and (5.24) in a Basic program for the Weibull failure distribution. Murthy and Iskandar (1992) used the same numerical method but expanded the program by including options for different failure distributions. Their program was implemented in Fortran and imposed a variable dimension limit of size 6000. Thus, for a step size h, the program could compute $M(t)$ for $t \leqslant 6000h$, which effectively means that time could be divided into as many as 6000 intervals. In the current study, these programs have been adapted into an S-Plus program, which is discussed in detail in the next chapter. Both Xie, and Murthy and Iskandar have shown that the method produces reasonably accurate results with a step size from 0.001 to 0.002 for t up to about 3.

An estimate of the variance of the expected number of renewals of a component can be obtained by working through the following steps:

1. Obtain a confidence region for the Weibull parameter set using (5.13);

2. Evaluate the number of renewals of the points along this confidence region using (5.25) and (5.24);

3. Obtain an estimate of the variance of the number of renewals using:

$$\hat{var}\left[M_p\left(t_w\right)\right] = \frac{\sum_{i=1}^{n}\left[M_{pi}\left(t_w\right) - \hat{M}_p\left(t_w\right)\right]^2}{n-1}, \tag{5.26}$$

where $M_{pi}\left(t_w\right)$ is the number of renewals of component p during the warranty period, t_w, at point i on the confidence region which contains n points, and $\hat{M}_p\left(t_w\right)$ is the estimated number of replacements of component p during the warranty period. The way in which this variance is then used to estimate a confidence interval for the cost of the warranty is discussed in the Section 5.8.

5.7 Modelling The Cost of Component Repair

The cost of repair is generally treated as a constant in the literature. This thesis extends warranty cost models by considering the cost of repair to be a random variable. This has been done because the manufacturer's data reveals that there is quite a range of costs in the repair of any single component. Examination of Figure 5.3 reveals that bimodal and tri-modal cost distributions are evident, as well as skewed distributions with long tails, and combinations of these distributions. Although some of these cost distributions could be modelled with two or three normal distributions, to model all these various densities individually for all components would be an unmanageable task. As the intent of this study is to obtain an overall estimate of the cost of the warranty, we have opted to simplify the modelling of the cost of repair by using one statistic for all components, and obtaining its variance. The choice of the statistic is discussed in subsequent subsection.

It may seem surprising that there is such a variation in the cost of repair of a component, since under the warranty, when a component fails, it is replaced by a new one. However, the warranty cost also includes a labour component. The manufacturer indicates a scale of charges that it expects to pay for a repair, based upon a known fault. This would be easier for the manufacturer to estimate in cases where previous repairs have been conducted. Known problems will have a set procedure formulated for the repair of the fault, whilst new problems could involve some exploration by the mechanic to resolve the prob-

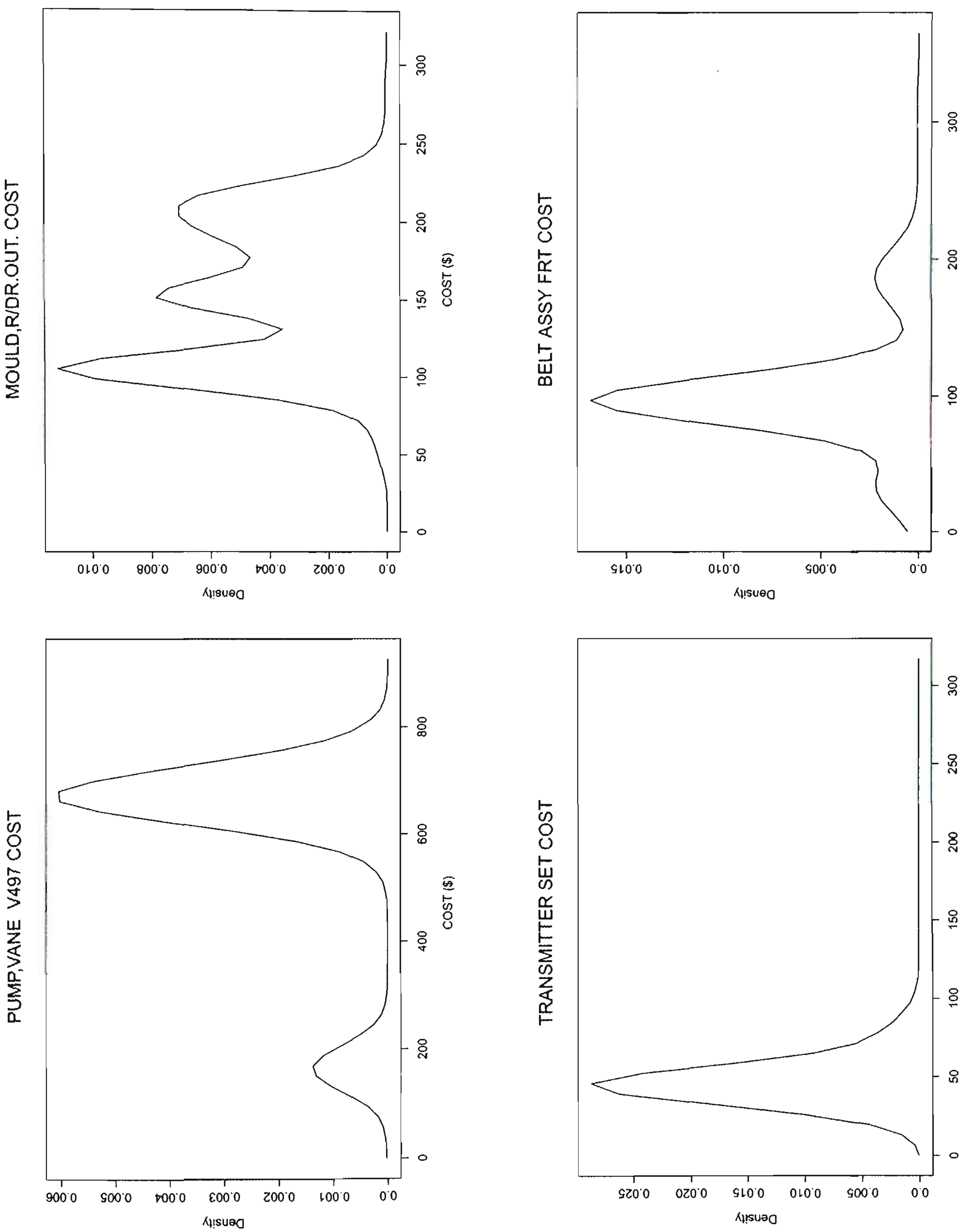


Fig. 5.3. *Cost density of some representative components.*

lem. As mechanics' skill levels vary, there would be a variation in the ability to diagnose and repair a problem from a customer's description. How well a customer can identify a problem, and how soon the customer acts upon a problem on first detecting it, could also influence the cost of repair. All of these factors can lead to a variation in the cost of repair for each component that is replaced, with a greater variation in more complex repairs.

Here is an example of this situation from the author's personal experience. A car owned by the author developed a loud knocking noise, which appeared to be coming from the transmission, and caused the shift lever to vibrate. The problem was actually caused by a faulty coil. This was well known to the repairer at the time, but could have been a bit of a mystery to a mechanic if he/she had not come across the problem before. In another example, a leak in the rear of the engine required the replacement of the rear main seal. This was a delicate operation, and if not fitted perfectly, the leak persisted, despite the replacement of the seal. When presented with the problem, one mechanic thought that the leak was coming from a different location, and replaced another seal, which of course, did not solve the problem. Eventually the vehicle was taken to another dealership, where there was a mechanic who was familiar with the problem and knew that the rear main seal needed replacing. He explained that it was difficult to fit the seal in exactly the right position when the engine was put back into place. Indeed, this mechanic needed two attempts to finally solve the problem. This example shows that mechanics have varying levels of skill and familiarity with a problem, and that an amount of exploratory work is sometimes required. Once a fault is well known, its solution should become more routine, take less time, and therefore be less costly to repair.

Lastly, it should be pointed out that some authors have built the value of money over time into the warranty cost model. Such studies include Hill and Blischke (1987), and Chun and Tang (1999). This is important when the time frame is quite large, or the inflation and investment rates are high. In our analysis, we are using a warranty of three years, which is a relatively short period of time. During our study, Australia has had a low inflation environment, as measured by changes in the Consumer Price Index. Over a four-year period, the overall inflation was less than 10% in total, with some quarters showing deflation. Another consideration is that the manufacturer collects the cost of the warranty when a vehicle is purchased, and is then able to invest the funds acquired. The increase in the cost of repair due to inflation should be nullified by the increase in investment capital. Thus, modelling the value of money in our models has been omitted.

However, an adjustment to the cost of the warranty to include the value of money can be incorporated into the models to obtain estimates in terms of today's dollar value.

5.7.1 Choice of a Representative Cost Statistic

From the discussion in the last subsection, it is clear that we are faced with choosing a statistic to represent the cost of repairing each component from a range of costs with various distributions. One option is the mean, which is efficient but not robust, and subject to the influence of outliers. Another option is the median, which is robust but not efficient. A compromise between the two, which is both efficient and robust, would appear to be a suitable choice. The trimmed mean, for example, is the average of the middle data (usually 50%). Thus, the influence of outliers is minimised at the expense of losing some data. With the trimmed mean, the smallest 25% of the data and largest 25% are not used, and only the inner 50% of the data are used. In contrast, the biweighted mean uses a more gradual weighting of the data (see Wonnacott and Wonnacott, 1990, page 539). The weighting, ω, on any data point X is

$$\omega = \begin{cases} (1 - Z^2)^2 & \text{if } |Z| \leqslant 1 \\ 0 & \text{if } |Z| \geqslant 1 \end{cases}, \tag{5.27}$$

where Z is a standardised X value given by

$$Z = \frac{X - \widetilde{X}}{3R}. \tag{5.28}$$

Here, $\widetilde{X}$ is the median and R is the interquartile range. Thus, the weighting decreases the further away the value of X is from the median, up to the point when X is three interquartile ranges or more from the median, when the weighting becomes zero. The biweighted mean is then given by

$$\bar{X}_b = \frac{\sum \omega X}{\sum \omega}. \tag{5.29}$$

A modification to this statistic is the biweighted iterated mean (see Wonnacott and Wonnacott, 1990, page 541), where the biweighted mean is used to again calculate the Z values. Thus, (5.28) is modified to

$$Z = \frac{X - \bar{X}_b}{3R}.$$

A set of weightings for the data points is then calculated according to (5.27) as before, and the biweighted iterated mean is calculated using (5.29). The process of recalculating Z values is repeated using the new value of $\bar{X}_b$, to obtain a new value of the biweighted

Table 5.1. *Comparison of median and various means.*

Part	Median	Mean	Trimmed Mean	Biweighed Mean	Biweighted Iterated Mean
BULB,LIC 12V	12.70	13.17	12.76	12.70	12.70
TRIMFIX	16.20	18.67	16.73	17.82	18.07
SOCKET& RET ASS	17.72	21.16	196.40	19.69	20.13
BULB LMP RH7802	18.80	19.73	18.61	18.06	17.91
BULB,RR LMP	19.81	20.84	19.90	19.86	19.88
RELAY,COMPUTER	39.34	38.30	39.21	39.28	39.26
TRANSMITTER SET	45.46	47.48	45.26	44.54	44.29
S/S	69.67	72.18	69.74	70.03	70.14
CAMRY BATTERY O	80.51	80.35	77.74	77.36	76.72
LAMP BACK UP	90.01	90.70	89.93	90.13	90.15
BELT ASSY FRT	95.88	102.38	96.85	96.58	96.67
MOULD.R/DR.OUT1	108.28	127.19	112.71	115.46	117.83
MOULD.F/DR.OUT1	121.13	136.87	121.28	118.85	118.28
MOULD.R/DR.OUT2	151.06	152.08	149.35	151.65	151.73
MOULD.F/DR.OUT2	171.62	174.15	172.27	173.55	173.83
DISC,FT	185.71	250.49	237.55	239.90	247.36
FRONT DOOR TRIM1	206.72	224.94	196.28	191.41	188.59
FRONT DOOR TRIM2	277.06	282.99	280.62	282.08	282.73
CARPET FLRFNTMC	329.29	327.49	336.77	341.21	342.99
PUMP,VANE	661.23	594.90	663.23	671.78	673.81

iterative mean. This process is repeated iteratively until the desired accuracy in $\bar{X}_b$ is achieved.

Thus, it would appear that any one of the trimmed mean, the biweighted mean or the biweighted iterative mean would be a good statistic to represent the cost of repairing a component. These values have been compared for several components and some of these are shown in Table 5.1. It appears, as the figures in the table reveal, that no one statistic seems to behave better than any other. It can be seen, for example, that the biweighted means lie outside the range set by the means and the medians for some components. Although the data for some components contain outliers, the mean has been the statistic used to represent the cost of repairing a component because the manufacturer's representative indicated that claims as large as the outliers were known to have been paid under the warranty. (This has been discussed in Section 4.3.) Thus, the mean cost of repair, $\bar{c}_p$, has been used as an estimate of the cost, c_p, of replacing component p. That is, an estimate of the cost of replacing component p is given by

$$\bar{c}_p = \frac{1}{r_p} \sum_{j=1}^{r_p} c_{p,j}, \tag{5.30}$$

where $c_{p,j}$ is the cost of replacing component p in the jth claim, and the sum is taken over all r_p claims involving component p.

5.7.2 Variance of the Cost of Repair

The variance of the cost of repairing component p can be estimated by the square of the standard deviation of the costs of repairing that component. That is,

$$\begin{aligned} \sigma_p^2 &= (sd_p)^2 \\ &= \frac{\sum_{p=1}^{r_p} (x_p - \bar{x}_p)^2}{r_p - 1}. \end{aligned} \tag{5.31}$$

where σ_p is the standard deviation of the (infinite) population of component p over unlimited time.

5.8 Modelling The Cost of a Warranty

The last two sections contain a discussion of how, on a component level, the number of repairs during the warranty period and the cost of those repairs can be estimated. Estimates of the variances of these quantities are included in the discussion. In this section, the point estimates and variances of the number of repairs and the cost of repairs are used to obtain an estimate of the warranty cost.

5.8.1 Point Estimate

For a warranty of duration t_w, the cost of the warranty per vehicle, w_p, on component p is given by

$$w_p = M_p(t_w)\, c_p, \tag{5.32}$$

where $M_p(t_w)$ is the expected number of replacements per vehicle of component p during the warranty, and c_p is the cost of the replacement. Note that the expected number of renewals, $M_p(t_w)$, during the warranty, and the cost of replacing component p, c_p, have been taken to be independent so that $cov\left[M_p(t_w), c_p\right] = 0$.

Equation (5.32) can be evaluated for the exponential model, where $M_p(t)$ is given by (5.20). In the Weibull model case, it can be found numerically using (5.25) and (5.24). An estimate of the cost, c_p, of replacing component p is obtained from (5.30).

5.8.2 Variance

The variance of a product of independent variables is given by

$$var\,(XY) = \mu_y^2\, var\,(X) + \mu_x^2\, var\,(Y) + var\,(X)\, var\,(Y)\,. \tag{5.33}$$

(See, for example, Mood, Graybill and Boes, 1963, page180.) Assuming that the number of replacements of a component is independent of the cost of replacing it, (5.33) holds.

Applying (5.33) to (5.32), the variance of the warranty cost, w_p, on component p is given by

$$var\,(w_p) = var\,[M_p\,(t_w) \times c_p]\,,$$

from which an estimate of the variance of the warranty cost of component p is

$$var\,(w_p) = \left[\hat{M}_p\,(t_w)\right]^2 var\,(c_p) + \bar{c}_p^2 var\,[M\,(t_w)] + var\,(c_p)\, var\,[M\,(t_w)]\,. \tag{5.34}$$

The expected number of renewals, $M_p\,(t_w)$, and its variance, $var\,[M\,(t_w)]$, can be evaluated for the exponential model using (5.20) and (5.21) respectively. An estimate of the cost of repair, c_p, and its variance, $var\,(c_p)$, can be obtained using (5.30) and (5.31). Thus, we get

$$\begin{aligned} var\,(w_p) &= (\lambda_p t_w)^2\, var\,(c_p) + \bar{c}_p^2 t_w^2 var\,(\lambda_p) + var\,(c_p)\, var\,(\lambda_p)\, t_w^2 \\ &= \left\{\lambda_p^2 var\,(c_p) + \bar{c}_p^2 var\,(\lambda_p) + var\,(c_p)\, var\,(\lambda_p)\right\} t_w^2. \end{aligned} \tag{5.35}$$

For the Weibull model, the expected number of renewals and its variance can be found numerically using (5.25) and (5.26) and the method discussed in Section 5.6.3. Again, (5.30) and (5.31) can be used to obtain estimates of c_p and $var\,(c_p)$ respectively.

5.8.3 Total Warranty Cost

Summing the point estimates over all components provides a point estimate of the overall warranty, W. That is

$$W = \sum_p w_p. \tag{5.36}$$

This can be evaluated using (5.32).

Using the fact that the variance of the sum of variables is the sum of the individual variances, we can obtain the variance of the total warranty cost as

$$var\,(W) = \sum_p var\,(w_p)\,. \tag{5.37}$$

This can be evaluated using (5.34), which can be evaluated for the exponential and Weibull models, as discussed in the last subsection.

To obtain a confidence interval for the warranty cost, a normal approximation can be used. Thus, a $100(1-2\alpha)\%$ confidence interval for the warranty cost is $W \pm z_\alpha \sqrt{var\,(W)}$, where z_α is the 100α percentile point of a standard normal distribution. Alternatively, the gamma distribution with shape parameter $= r$ and rate parameter $= \lambda$, such that $E\,(X) = \frac{r}{\lambda}$ and $\text{var}(X) = \frac{r}{\lambda^2}$ can be used. The results from these two methods are compared in Section 6.7.

5.8.4 Point and Interval Estimates Using the Bootstrap Method

Background

A less traditional approach in obtaining a point and interval estimate of the total cost of a warranty is the bootstrap method. This technique can be used to obtain a point estimate and a confidence interval for any statistic, such as the mean, the median, the standard deviation, or indeed a function of the variables. The method uses sampling with replacement, which is repeated a large number of times, say, 1000. Inferences can then be made from the distribution of these samples. Details of the bootstrap theory can be found in Efron and Tibshirani (1993). Following is a précis of some of the theory that has been used in this warranty application.

An estimate of the distribution of any statistic, in our case the warranty cost per vehicle, can be obtained from the distribution of that statistic obtained from the sampling. For a set of observed data points $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, a bootstrap sample is obtained by random sampling n times, with replacement, from the observed data, and is denoted by $\mathbf{x}^* = (x_1^*, x_2^*, \ldots, x_n^*)$. A large number, B, of independent bootstrap samples, $\mathbf{x}^{*1}, \mathbf{x}^{*2}, \ldots, \mathbf{x}^{*B}$ are then generated. For the desired statistic, $s\,(\mathbf{x})$, such as the mean, a bootstrap replication of s, namely $s\left(\mathbf{x}^{*b}\right)$, is obtained for each sample. The bootstrap estimate of the standard error is the standard deviation of the bootstrap replications,

$$\hat{s}_{boot} = \left\{ \frac{\sum_{b=1}^{B} \left[s\left(\mathbf{x}^{*b}\right) - s\,(\cdot) \right]^2}{B-1} \right\}^{\frac{1}{2}}, \tag{5.38}$$

where $s\,(\cdot) = \sum_{b=1}^{B} s\left(\mathbf{x}^{*b}\right) / B$.

A confidence interval for $s(\mathbf{x})$ can be obtained from the percentile interval of the distribution of the value of this statistic, $s\left(\mathbf{x}^{*b}\right)$, obtained from the bootstrap samples. Let $\hat{\theta}^{*b} = s\left(\mathbf{x}^{*b}\right)$ be the bootstrap estimate of the required statistic for the bootstrap sample b. Then the $100(1-2\alpha)\%$ confidence interval for that statistic can be defined by the α and $1-\alpha$ percentiles of the distribution of $\hat{\theta}^{*}$. This may be written as

$$\left[\hat{\theta}_l, \hat{\theta}_u\right] = \left[\hat{\theta}^{*(\alpha)}, \hat{\theta}^{*(1-\alpha)}\right]. \tag{5.39}$$

By the central limit theorem, as $n \to \infty$, the distribution of the bootstrap statistic, $\hat{\theta}^{*}$, will become normal. Further, for small n, the confidence interval defined in (5.39) has the following desirable properties: it has good coverage, that is, a proportion α of the $\hat{\theta}^{*b}$ lie below this range, with the same proportion lying above; it is transformation-respecting, that is, the percentile interval for any monotone transformation $\phi = m(\theta)$ is the percentile interval for θ mapped by $m(\theta)$; and it is range-preserving, that is, any restriction on the range of values that a parameter can have is also imposed on the bootstrap statistic $\hat{\theta}^{*}$.

An improvement on the percentile confidence interval is the BC_a (bias-corrected and accelerated) interval. The BC_a interval corrects for bias in the percentile interval. It is given by

$$\left[\hat{\theta}_l, \hat{\theta}_u\right] = \left[\hat{\theta}^{*(\alpha_1)}, \hat{\theta}^{*(\alpha_2)}\right], \tag{5.40}$$

where

$$\begin{aligned} \alpha_1 &= \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(\alpha)}}{1 - \hat{a}\left(\hat{z}_0 + z^{(\alpha)}\right)}\right) \\ \alpha_2 &= \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(1-\alpha)}}{1 - \hat{a}\left(\hat{z}_0 + z^{(1-\alpha)}\right)}\right). \end{aligned} \tag{5.41}$$

$\Phi(\cdot)$ is the cumulative normal distribution function and $z^{(\alpha)}$ is the 100α percentile point of a standard normal distribution. The values of α_1 and α_2 depend on the values of $\hat{z}_0$ and $\hat{a}$.

The value of the bias-correction, $\hat{z}_0$, is obtained from the proportion of bootstrap replications less than the original estimate $\hat{\theta}$,

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\#\left\{\hat{\theta}^{*b} < \hat{\theta}\right\}}{B}\right), \tag{5.42}$$

where $\Phi^{-1}(\cdot)$ is the inverse function of a standard normal cumulative distribution function, and $\#\{\cdot\}$ refers to the number of elements in a set.

The acceleration, $\hat{a}$, is calculated in terms of the jackknife values of a statistic $\hat{\theta} = s(\mathbf{x})$. Let $\mathbf{x}_{(i)}$ be the original sample with the ith point x_i removed, and let $\hat{\theta}_{(i)} = s\left(\mathbf{x}_{(i)}\right)$. Define $\hat{\theta}_{(\cdot)} = \sum_{i=1}^{n} \hat{\theta}_{(i)}/n$. Then an expression for the acceleration is

$$\hat{a} = \frac{\sum_{i=1}^{n}\left(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)}\right)^3}{6\left\{\sum_{i=1}^{n}\left(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)}\right)^2\right\}^{3/2}}. \quad (5.43)$$

Warranty Application

The warranty cost per vehicle, $\hat{W}$, can be obtained by generating a number of bootstrap samples, obtaining this cost for each sample, and then obtaining the mean warranty cost per vehicle of all the samples. However, a more direct approach is to obtain an estimate of this statistic from the observed sample of vehicles with claims. Thus we have

$$\hat{W} = \sum_{j=1}^{r} \frac{c_j}{n},$$

where c_j is the cost of repairing vehicle j, and r is the number of claims made during the warranty period. Alternatively, $\hat{W}$ can be obtained by considering the expected warranty cost per vehicle, and summing over all components, as in (5.36).

An estimate of the confidence interval on the cost of a warranty can be obtained from the BC_a confidence interval. Using (5.40), (5.41), (5.42) and (5.43), a BC_a confidence interval can be constructed. However, a BC_a interval can be readily evaluated in most statistical packages.

5.9 Limitations of the Models

It should be pointed out that the above modelling is subject to the assumptions discussed at the beginning of this chapter in Section 5.2.

Moreover, in the previous sections, only the exponential and Weibull distributions have been discussed to model failure of components. Because of the vast range of components that go into making up a vehicle, it may be inaccurate to model the reliability of all components using just these two distributions. However, as discussed in Section 6.3, there is not much difference in the log likelihood values amongst the various alternative models that have been considered. The Weibull model provides a good fit in the majority of cases. It is a versatile model as it encompasses other models. For this reason, it is one of the preferred models.

The analysis in this study has been based on time only. Since the vehicle warranty here is two-dimensional, the modelling that has been investigated is a simplification of reality. This has been discussed in Section 5.2.

It also needs to be pointed out that extrapolating the models much beyond the warranty period could lead to erroneous predictions. This is because the parameters of the models are based on observations made during the warranty period, which allows for the observation of only the early failures of components, compared to mean life of most components.

5.10 Conclusion

This chapter has discussed models for the costing of a warranty. Identifying the assumptions made in establishing the models was the first step of the process. Survival analysis techniques, which take into consideration the fact that not all components are observed until failure, have been used in developing the models. Both parametric and non-parametric models have been used. The cost models used in this chapter have been based on the exponential and Weibull only. However, several other models have been considered, many of which are encompassed within the Weibull model. Renewal theory has been called upon to estimate the expected number of claims during the warranty period. Both a point estimate of the number of renewals and an estimate of its standard error have been obtained. The costs of repairing components have been taken to be random variables, rather than fixed amounts, so that both point estimates and standard deviations have been obtained. Using the point estimates of the number of renewals and the cost of repairing each component, point and interval estimates of the warranty cost due to each component have been obtained, from which the total cost of a warranty has been established.

This approach has been complemented by the bootstrap approach of obtaining an estimate of the warranty cost by resampling from the observed data. This approach has been far simpler to implement, as it does not rely on the analysis of each component individually. However, if more detail on individual components, or extrapolation beyond the present limits of the data, is required, then the former approach is more appropriate.

In the next chapter, the models discussed in this chapter are applied to the manufacturer's data. Problems involved in dealing with the real data, and how they have been overcome are discussed.

Chapter 6
Implementation of the Survival Models

6.1 Introduction

The last chapter looked at the theory behind the survival models. This chapter discusses the application of these models to the manufacturer's database. The purpose of this chapter is to obtain the warranty cost and variance per vehicle manufactured, firstly for each component, and then for the entire vehicle, using data from a whole year's production.

Preliminary data analysis of the manufacturer's database has revealed that a number of claims were paid after the warranty period. In cases of known manufacturing faults, the manufacturer honoured claims beyond the warranty period. However, the number of claims made after the warranty expired is not representative of the total number of failures at that time, as most owners would, at that point, pay for the repairs rather than make a warranty claim. After the warranty, the vast majority of vehicles are no longer being observed (that is, they become censored items), which greatly reduces the number of vehicles at risk. Therefore, an estimate of the probability of failure takes an enormous jump because of the reduced sample size. Thus, no claims beyond three years have been used to model the failure of components.

Section 4.6 discussed how a survival database containing the fields VIN, $age2$, $dist2$, $part$, $partname$ and $status$ had been created after cleaning the data. The database is in a suitable format to fit survival models for each part, which is the subject of this chapter. Thus, the survival models discussed in the previous chapter are implemented in this chapter. Even though the previous chapter only discussed the exponential and Weibull parametric models, this chapter looks at various other models which are compared to these two models.

This chapter starts by examining the distribution of the number of claims by component, and then discusses appropriate analysis for the number of claims being made. A look at possible model candidates precedes a discussion of the exponential and Weibull models of component reliability. As in the last chapter, the reliability models for each component are used in conjunction with renewal theory to obtain point estimates and standard errors of the number of renewals during the warranty period. The repair cost models developed in the last chapter are also applied to the manufacturer's data. The individual component warranty-cost models that have been developed in the last chapter are then applied to the manufacturer's data. Using the model parameters obtained from the manufacturer's data, the cost of extending the warranty is estimated. The chapter ends with a test of the sensitivity to change of the warranty cost when the Weibull parameter estimates vary from their determined values.

As in the previous chapters, a number of original S-Plus functions and scripts have been written by the author for the analysis required in this chapter. Many are presented in Appendix C. All the functions, scripts and data frames referred to in this chapter can be viewed on the accompanying compact disc.

6.2 Frequency Distribution of Claims

The purpose of this section is to examine the distribution of the number of claims. The number of different types of component that exist in the manufacturer's database, and the number of claims on each of these component types, are investigated. The distributions of the number of claims is then studied. It should be restated here that once a component fails, it is replaced with a new one. This new one could itself fail before the vehicle it is installed in reaches the end of its warranty. Thus, more than one claim per vehicle on a component type is possible during its warranty period.

The manufacturer's database covers various configurations of one make and model of a passenger vehicle. The variants include different engines, body configurations, and equipment levels. Components that were provided by different suppliers have been identified by different part numbers. Thus, a particular part-specification could have several different part numbers. These have all been analysed separately. Thus, it has not been assumed that parts supplied by different vendors have identical reliability distributions. Therefore, although 3,040 different parts, as identified by unique part numbers, have been identified in the manufacturer's database, no one vehicle contains that many components.

The vast majority of parts in the manufacturer's database proved to be very reliable during the warranty period. Of the 59,934 claims made during the warranty, there were

Table 6.1. *Number of parts with given number of claims.*

(a) All Claims.

Number of Claims	Number of part types
1-100	2964
101-200	43
201-300	12
301-400	5
401-500	3
501-600	3
601-700	1
701-800	2
801-900	2
900-1000	0
>1000	6

(b) Up to 100 claims.

Number of Claims	Number of part types
1-10	2498
11-20	207
21-30	91
31-40	58
41-50	45
51-60	24
61-70	11
71-80	13
81-90	8
91-100	9

(c) Up to 10 claims.

Number of Claims	Number of part types
1	1159
2	454
3	262
4	193
5	141
6	89
7	63
8	62
9	44
10	31

1,159 parts that were replaced once only. Thus, it has only been possible to perform limited reliability analysis on these parts. Table 6.1(a) shows the distribution of the number of different parts by the number of claims. The 'number of claims' refers to the number of times a warranty payment has been made on a particular type of component. It is clear from the table that 100 claims or fewer were made against the vast majority of parts.

Table 6.1(b) shows a further breakdown of the number of parts with up to 100 claims. This table again reveals that the vast majority of parts had very few claims made against them. It can be seen, for example, that there were 10 or fewer claims made on 2,498 different parts.

Table 6.1(c) shows yet a further breakdown of the number of parts with up to 10 claims. It shows how few claims most parts had made against them. For example, 1,159 parts had only one claim made against them during the warranty period, whilst 454 parts had two claims made against them.

The above analysis was conducted in S-Plus using the script *PartFreq.ssc*, which is included on the accompanying compact disc.

It is reasonable to use a simple model to describe the reliability of parts with few claims. An exponential model, which is based on a constant failure rate, is appropriate in this case.

We shall refer to components with 1, 2, 3, . . . claims occurring during the warranty period as one-claim, two-claim, three-claim, . . . components.

6.3 Parametric Models of Reliability

A number of distributions have been considered to model the reliability of the individual components. They are the extreme, normal, logistic, Rayleigh, lognormal and loglogistic distributions. They have been selected not only because they are readily available in the S-Plus survival analysis function *survReg*, but also because these distributions have been used to various extents in the reliability literature previously.

No one model was found to provide the best fit for all components. In fact, for most components all models fit the data reasonably well, with little difference between their log likelihood values. More than any other model, the lognormal, followed by the Weibull, have provided the best fit. However, the lognormal model proved to be unreliable for several components, with its log likelihood values being several orders of magnitude larger than that of the other models. The Weibull model appears to be the most suitable for components with a reasonable number of claims, as it is the most versatile. It encompasses many other distributions (Nelson, 1982, page 37). If $\beta = 1$, the Weibull simplifies to the exponential distribution, whilst if $\beta = 2$, it becomes the Rayleigh distribution. If $3 \leqslant \beta \leqslant 4$, the Weibull distribution is close to the normal distribution. For large values of β, say $\beta \geqslant 10$, the shape of the Weibull distribution is close to that of the smallest extreme-value distribution. Nelson suggested that the Weibull distribution is the most suitable distribution for life data. For components with fewer claims, the exponential distribution is suitable to model reliability because of its mathematical tractability.

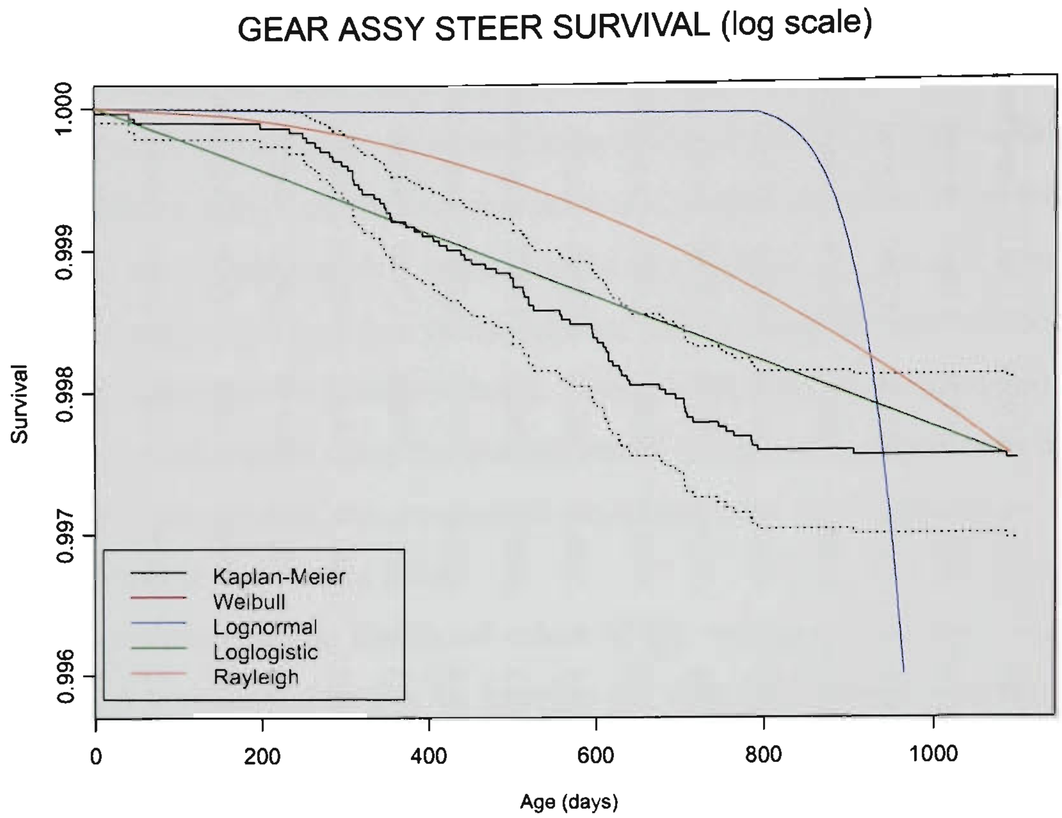
As discussed in the last section, for the majority of parts, few claims have been made. From Table 6.1(b), it can be seen that there is somewhat of a drop in the frequency in the number of components with more than twenty claims. By using the exponential model for components less than twenty claims, before using the Weibull model, the amount of processing can be greatly reduced. Thus, twenty claims is considered to be a suitable cut-off point to start using the Weibull model in this study. In order to compare the results of the two models, the exponential model has been used to model all components, as discussed in Section 6.4 below.

Table 6.2 shows the log likelihood values of the various models for a sample of components. A complete table can be found in the data frame *logLike*, which can be opened through the object explorer in an S-Plus session. The table show that in most instances the log likelihood values of the models do not vary greatly. However, there are some notable exceptions. For example, the log likelihoods of the lognormal models for components "GEAR ASSY P SXV" and "GEAR ASSY STEER" are several orders of magnitude smaller than the log likelihoods of the other models for the respective parts. Similarly, the extreme model's log likelihood value of the component "TRANSMITTER SET" is also vastly different to the other models' values. This puts these two particular models at odds with all the other models. Table 6.2 and Figure 6.1 clearly show the discrepancy with the other models for the latter two components. Log likelihood values that are several orders of magnitude lower than the other models' respective log likelihood values have occurred with only the extreme, normal and lognormal models. The reason for this has not been pursued. As the Weibull model is the most universal, it is the preferred model for components with twenty or more claims. The Weibull model's log likelihood values are not necessarily the largest, but they are of the same order of magnitude as the others.

Of the components with twenty or more claims, the Weibull model does not fit fourteen of the 347 components particularly well. This can be seen by the low log likelihood values of the Weibull model compared with the other models' log likelihood values. The fact that the Weibull model does not fit these fourteen components well can also be seen by examining the survival graphs of these components in Appendix D.1. It has been dis-

Table 6.2. *Log likelihood of various models for some components.*

Part	No. of Claims	Extreme	Normal	Logistic	Exponential	Weibull	Lognormal	Loglogistic	Rayleigh	Beta	Best Fit
KIT DRIVE SHAFT	5	-84.117	-83.998	-84.117	-83.522	-82.901	-82.794	-82.9009	-82.9599	1.72212	Lognormal
GEAR ASSY P SXV	10	-162.44	-162.27	-162.44	-160.11	-160.10	$-6.805x10^{38}$	-160.10	-163.71	0.95016	Weibull
GSKT,CYL HD NO2	15	-231.87	-231.75	-231.87	-234.09	-232.29	-2789.6	-232.29	-232.50	1.7030	Normal
INSULATOR, EXHA	20	-300.27	-299.90	-300.26	-306.36	-298.42	-298.18	-298.42	-299.53	2.8454	Lognormal
RAD HOSE UPPER	30	-423.47	-423.26	-423.47	-447.38	-423.15	-423.33	-423.15	-432.60	4.9895	Weibull
W/STRIP DOOR	50	-719.76	-718.57	-719.76	-720.09	-711.98	-711.39	-711.98	-712.08	1.8766	Lognormal
GEAR ASSY STEER	75	-1079.9	-1077.1	-1079.9	-1049.7	-1049.70	$-3.402x10^{38}$	-1049.7	-1070.7	1.0284	Loglogistic
4SPK RAD/CASS	102	-1425.2	-1422.3	-1425.2	-1396.3	-1396.3	-1397.5	-1396.3	-1427.7	0.9990	Weibull
DISC FRONT	152	-2041.5	-2036.6	-2041.4	-2020.1	-2009.7	-2010.3	-2009.7	-2017.3	1.4828	Loglogistic
NEW VEHICLE SPE	200	-2789.5	-2779.4	-2789.4	-2603.1	-2224.0	-2216.6	-2223.8	-3339.2	0.2291	Lognormal
PUMP,VANE V497	302	-3772.7	-3766.7	-3772.6	-3806.2	-3765.2	-3787.3	-3765.3	-3767.6	1.7670	Weibull
BULB LMP RH7802	499	-5866.7	-5870.9	-5866.8	-6038.5	-6000.0	-6099.8	-6001.4	-6020.7	1.5205	Extreme
MOULD.R/DR.OUT.	705	-8169.9	-8150.1	-8169.2	-8287.8	-8132.9	-8168.6	-8133.0	-8133.6	2.0949	Weibull
TRANSMITTER SET	1662	$-3.402x10^{38}$	-18135	-18196	-18113	-17953	-18006	-17952	-18001	1.5929	Loglogistic
MOULD F/DR.OUT.	2035	-21322	-21277	-21318	-21766	-21390	-21712	-21399	-21391	1.9405	Normal
TRIMFIX	2964	-31594	-31426	-31573	-30588	-30355	-30638	-30377	-32899	0.6938	Weibull
TRANSMIT CTRL S	8733	-84449	-83648	-84227	-80687	-80684	-80636	-80625	-84014	0.9751	Loglogistic}



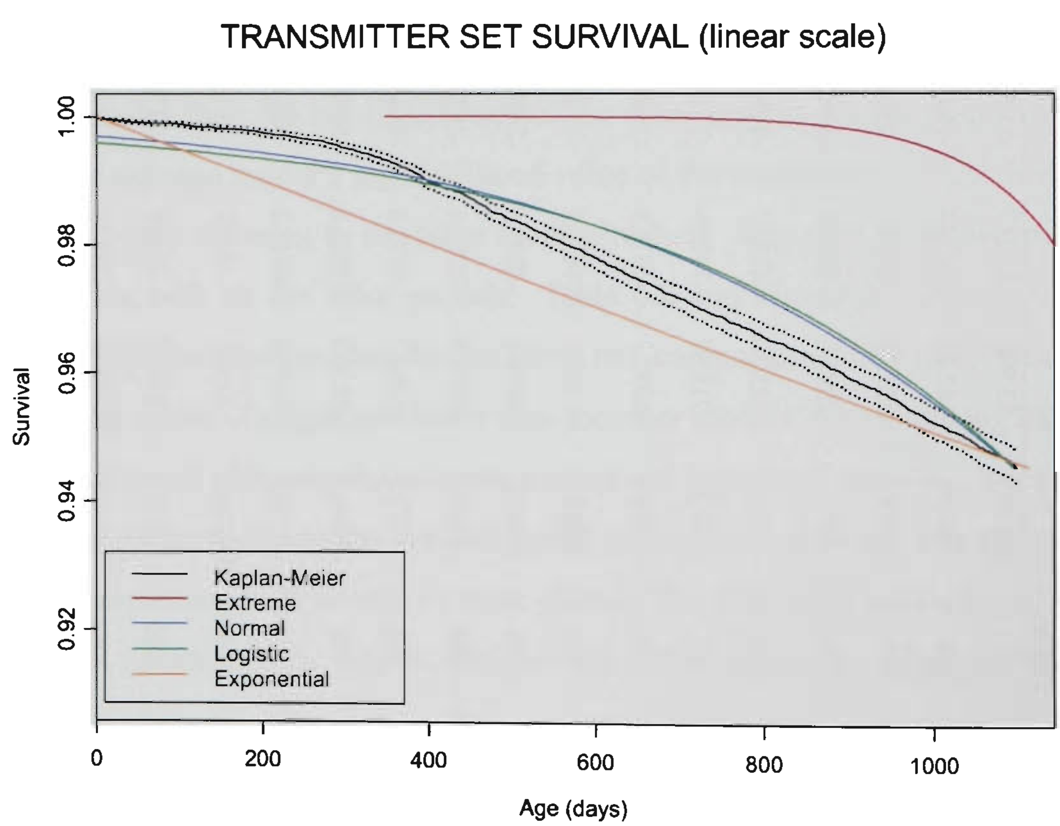


Fig. 6.1. *Examples of ill-fitting models.*

covered that a better fit could be obtained by removing one or more points with the highest values of $age2$, that is, the points representing the oldest non-censored components. In all but three cases of the fourteen, only one point has to be removed in order to obtain a closely-fitting Weibull model. In two cases, two points had to be trimmed, and in one case, three points, in order to obtain a closely-fitting Weibull model. These results can be seen in Appendix D.1. In all fourteen cases, the removal of these points had resulted in a Weibull model with a higher log likelihood value than the corresponding exponential model's log likelihood. Although the ensuing Weibull model does not always produce the highest log likelihood values, they are very close to the maximum values amongst the considered models.

There has not appeared to be any significant difference in the distributions of $age2$ between components with an initially poor Weibull fit, and those with a good Weibull fit. This can be seen by studying the figures in Appendices D.2 and D.3, which show the distributions of $age2$ of a sample of components from each of the two groups. Indeed, the only difference in the distribution of $age2$ before and after the removal of largest value(s) is the absence of the removed point(s). It is unclear as to why the removal of points with the largest values of $age2$ lead to a much better Weibull fit. Although this observation is worthy of further investigation, it was not possible to pursue it in the current study.

6.4 Exponential Model

This section looks at the modelling of warranty costs using the exponential distribution. As previously discussed in Section 5.5, the exponential model is appropriate when there are few failures in a large sample. The resulting survival data will contain mainly censored items. All components, not just the ones with a low failure rate, have been analysed using the exponential model in order to compare it with the Weibull model.

This section begins with a discussion of a point estimate of the cost of a warranty when the reliability of components is modelled by the exponential distribution. This is followed by a discussion of the variance of this estimate. The method of obtaining a point estimate and variance using S-Plus is discussed in the section after that. Following that, a simplified implementation is presented. The section ends with a discussion of the results of the two different approaches.

6.4.1 Expected Warranty Cost of Components

Point Estimate

Equation (5.32) gives the warranty cost, w_p of component p as

$$w_p = M_p\left(t_w\right) c_p,$$

where $M_p\left(t_w\right)$ is the expected number of replacements of component p during the warranty period t_w, and c_p is the cost of the replacement. The expected number of replacements can be evaluated using (5.20). That is,

$$M_p\left(t_w\right) = \lambda_p t_w,$$

An estimate of c_p can be given by the mean cost in (5.30) as

$$\bar{c}_p = \frac{1}{r_p}\sum_{j=1}^{r_p} c_{p,j},$$

where r_p is the number of claims involving component p. Thus, the warranty cost of component p is

$$w_p = \frac{\lambda_p t_w}{r_p}\sum_{j=1}^{r_p} c_{p,j}. \tag{6.1}$$

The warranty cost per vehicle over all components is given in by (5.36). Taking the sum over all components to be modelled by the exponential distribution and using (6.1), we get the total warranty cost of all components being modelled by the exponential distribution as

$$W_E = t_w \sum_{p\in E}\frac{\lambda_p}{r_p}\sum_{j=1}^{r_p} c_{p,j}, \tag{6.2}$$

where E is the set of components being modelled exponentially.

Variance

An expression for the variance of the warranty cost of component p is given by (5.35), which is,

$$var\left(w_p\right) = \left\{\lambda_p^2 var\left(c_p\right) + \bar{c}_p^2 var\left(\lambda_p\right) + var\left(c_p\right) var\left(\lambda_p\right)\right\} t_w^2.$$

The mean cost of repair, $\bar{c}_p$, and the variance of this cost, $var\left(c_p\right)$, can be obtained using (5.30) and (5.31). By (5.37), the variance of the total warranty cost of the components being modelled by the exponential distribution is the sum of the individual variances. That is,

$$var\left(W\right) = \sum_{p} var\left(w_p\right), \tag{6.3}$$

where the sum is taken over all components being modelled exponentially.

Once an estimate of the total warranty cost, W, and its variance, $var\left(W\right)$, are obtained, a confidence interval of the cost of the warranty can be calculated. An approximate $(1 - 2\alpha) \times 100\%$ two-sided confidence interval is provided by the expression

$$W \pm z_{\alpha} var\left(W\right),$$

where z_a is the standard normal value having a right tail area of α.

6.4.2 Implementation in S-Plus

The S-Plus function *fCstExp* implements (6.1) and (5.35), and it is included in Appendix C.1. The inputs to the function include a data frame of component failure times and a range of number of claims identifying which components are to be analysed. The function returns a data frame containing each component's average repair cost and the standard deviation of this cost. The function also fits an exponential model to each component and returns both the failure rate and the standard deviation of the failure rate. Lastly, the function computes the warranty cost of each component together with the variance and standard deviation of this cost. Since the variance of the cost of repair of one-claim components cannot be calculated because there is only one observation, the cost of repair has been taken as a constant for these components and its variance has been taken to be zero.

In the *fCstExp* function, the unit of time is days. Consequently, the failure rate is the number of failures per day. Although this unit is somewhat awkward, it has been used so that a comparison with the results of the Weibull modelling can be made.

As function *fCstExp* is processor intensive, it is unable to process 2,000 records at once. The processor (in this case and AMD1800 with 512MB SDRAM) runs out of dynamic memory. In order to process all the components, it is necessary to break up the task into smaller jobs, and then combine the results. This has been done in the script *RunCostExp.ssc*, which can be found on the accompanying compact disc. The results of the calculations are discussed in Section 6.4.3.

Simplified Implementation

Since the failure rate does not vary greatly between components with the same number of claims, a simplified analysis involves modelling such components together, with the cost of repair being the sum of the average costs of repairing each component. Consequently, all one-claim components can be analysed simultaneously. Likewise, all two-claim, and then three-claim components, etcetera, can also be modelled together in groups.

In order to investigate the differences in the failure rates of components with failures occurring at different times during the warranty period, the function *fFailRate* has been written (see Appendix C.2). Three scenarios of component failures are simulated in the function:

- All failures occur at the start of the warranty period;
- All failures occur at the end of the warranty period;
- Failures are evenly distributed throughout the warranty period.

It was established that the timing of the failures has little impact on the failure rates.

The script *ExpRteCmpr.ssc*, which is on the accompanying compact disc, was developed to compare the results of the grouped approach with the individual analysis approach for components with up to fifty claims. In the script, for each k-claims group, the individual component's failure rates are compared with the mean of the group, and with the rates obtained from the above three scenarios. The results are in the data frame *expRteCmpr* on the accompanying compact disc, and the conclusions that have been drawn are:

- For each k-claims group, the minimum and maximum failure rates are within 0.05% of the mean rate of the group.
- The mean failure rates of each k-claims group are within 0.001% of the rate when failures are evenly distributed throughout the warranty period.
- The failure rates when all claims are made at the start or end of the warranty period are within 0.01% of the rate when the failures are evenly distributed throughout the warranty period.

Thus, it can be concluded that for the simplified analysis, with little loss of accuracy, is the preferable approach.

The function *fCstExpGp* uses the grouped-claims approach (see Appendix C.3). This somewhat reduces the processing time, as all components with the same number of claims are analysed together. For example, the 1,159 one-claim components can be analysed as one group, with one calculation, as opposed to being analysed individually, which would require 1,159 calculations.

The warranty cost obtained by the use of individual-component analysis was compared with the warranty cost obtained by the grouping of k-claims components together. Considering components with 50 or less claims, the difference in total warranty cost calculated by the two different methods was 0.05% of the total warranty cost of individual components. Similarly, the difference in the variance provided by the two methods was 0.09% of the total variance of individual components. Thus, it can be concluded that it is more expedient to obtain an estimate of the warranty cost by grouping all k-claims components and analysing them as one group.

6.4.3 Results

A table of results from the individual-components analysis can be found in the data frame *costExp* on the accompanying compact disc. The results of the analysis involving the grouping of k-claims components can be found in the data set *cstExpGp* on the compact disc.

Table 6.3 shows the comparison of the warranty cost per vehicle, and the standard deviation of this cost, using the two different analyses for components with up to fifty claims. The frequency of components with more than fifty claims is low, and in most cases there is only one component with a particular frequency. Thus, little is to be gained by grouping the components into same-frequency categories. The table shows that the two different methods produce very close results for both the warranty cost per vehicle and the standard deviation of this cost.

For those components having up to 50 claims, the total warranty cost per vehicle and variance of this cost, using the two different methods, are shown in Table 6.4. It can be seen that the two methods provide similar results. Later in Section 6.6, we have compared the results obtained from modelling the total warranty cost using the individual-

Table 6.3. *Comparison of individual and grouped warranty costs and standard deviations.*

Number of Claims	Individual Cost	Grouped Cost	Individual Std Dev	Group Std Dev
1	12.00711	12.00669	1.43517	1.43513
2	8.38868	8.38813	0.86184	0.86179
3	6.56014	6.55949	0.76708	0.76700
4	5.01652	5.01585	0.63794	0.63785
5	3.87094	3.87030	0.49984	0.49976
6	3.34195	3.34129	0.92390	0.92371
7	1.87243	1.87200	0.20468	0.20464
8	3.44628	3.44536	0.56891	0.56876
9	2.91179	2.91092	0.38423	0.38411
10	1.55199	1.55147	0.21568	0.21561
11	2.54064	2.53972	0.48264	0.48246
12	2.64094	2.63989	0.38336	0.38321
13	1.63462	1.63392	0.24382	0.24372
14	3.47424	3.47262	0.50901	0.50877
15	1.80178	1.80089	0.26220	0.26207
16	0.88828	0.88781	0.17734	0.17725
17	3.98320	3.98096	0.57878	0.57845
18	1.08326	1.08261	0.25316	0.25301
19	1.33202	1.33119	0.22897	0.22882
20	0.82204	0.82150	0.14990	0.14980
21	1.77583	1.77459	0.31875	0.31853
22	0.94539	0.94470	0.17140	0.17128
23	1.78654	1.78518	0.24291	0.24272
24	1.30345	1.30241	0.27178	0.27156
25	0.74982	0.74920	0.16120	0.16107
26	0.55613	0.55565	0.10067	0.10058
27	0.37103	0.37069	0.12657	0.12646
28	0.62016	0.61959	0.16520	0.16505
29	0.37830	0.37794	0.11972	0.11960
30	1.68519	1.68351	0.21931	0.21909
31	1.27277	1.27146	0.17709	0.17690
32	0.92457	0.92359	0.19422	0.19402
33	1.49763	1.49599	0.57352	0.57289
34	0.70032	0.69953	0.21126	0.21102
35	1.45376	1.45208	0.19990	0.19967
36	2.06584	2.06340	0.60075	0.60004
37	0.80630	0.80531	0.17554	0.17533
38	1.61239	1.61036	0.45554	0.45497
39	1.25653	1.25488	0.31445	0.31404
40	0.75721	0.75620	0.14081	0.14062
41	0.95832	0.95702	0.25527	0.25492
42	0.38258	0.38205	0.07553	0.07542
43	0.70621	0.70520	0.25625	0.25588
44	0.86015	0.85890	0.20806	0.20775
45	0.49257	0.49184	0.19319	0.19290
46	1.33501	1.33298	0.21441	0.21408
47	1.11428	1.11255	0.32034	0.31984
48	0.70865	0.70753	0.37523	0.37463
49	1.48290	1.48049	0.18422	0.18392
50	0.42345	0.42275	0.15921	0.15894

Table 6.4. *Warranty cost and variance due to components having up to 50 claims.*

	Individual	Grouped
Cost	100.1521	100.1001
Variance	8.905666	8.898051

component and the grouped-component exponential models, with the results obtained from the Weibull model.

Thus, good results for both warranty cost and variance were obtained by modelling the failure of components either individually or in groups. It was shown that the results for the warranty cost and variance obtained by either method are similar.

6.5 Weibull Model

This section discusses the development of a warranty cost model using the Weibull distribution. It begins with a lengthy discussion of the S-Plus function *survReg*, which is used to fit a parametric model to survival data. This subsection has been included because the output of the function needs some interpretation. Following this is a discussion of how the warranty cost models developed in the previous chapter are implemented in S-Plus to obtain a point estimate and its variance. The section ends with a discourse on the results of the analysis.

6.5.1 The S-Plus Function *survReg*

The standard S-Plus function *survReg* has been used to fit a model to a given set of survival data. Before beginning the Weibull analysis, an interpretation of the meaning of the output of the *survReg* function, both in the S-Plus manuals and in the wider literature, has been sought. There appears to be a lack of clear documentation in either source. We have endeavoured to deduce the meaning of the *survReg* function output by interpreting references to parametric survival models in the S-Plus manuals and in Venables and Ripley (1999). An examination of the contents of the section on linear modelling in the S-Plus manuals has also been undertaken. The following paragraphs discusses the findings.

Firstly, *survReg* uses maximum likelihood techniques to obtain estimates of the parameters of the distribution used to model the data. The extreme-value, normal and logistic distributions, together with their log counterparts, the Weibull, lognormal, loglogistic, as well as the exponential and Rayleigh distributions are available in *survReg*.

The S-Plus 2000 Guide to Statistics, Volume 2 (page 348), in its Chapter "Parametric Regression in Survival Models" states that the Weibull distribution has a hazard function of $p\lambda\left(\lambda t\right)^{p-1}$. This implies that the survival function for the Weibull distribution is $S\left(t\right) = \exp\left(-\left(\lambda t\right)^{p}\right)$. This is equivalent to (5.8) with $p = \beta$ and $\lambda = 1/\eta$. It is not made clear in the S-Plus documentation how these parameters relate to the output of the *survReg* function. Assigning fit to the output of *survReg*, the components of the output that relate to the Weibull parameters are $fit\$coeff$ and *fit$scale*.

In the section on Parametric Survival Models in the S-Plus 2000 Guide to Statistics, Volume 2 (page 370 ff.), a link function $g(\cdot)$ is used to identify a specific distribution from a family of distributions. Let

$$z = \frac{g\left(y\right) - x\beta}{\sigma} \tag{6.4}$$

be the random variable for the failure time y, where σ is the scale factor, x is a vector of co-variates, and β is a vector of coefficients. In our situation, we have $x = 1$, as there are co-variates. If we put $g\left(x\right) = \log x$ (referred to as the "log" link), we obtain the Weibull distribution from the smallest extreme-value distribution, the latter having a density given by

$$f\left(z\right) = \exp\left\{z - \exp\left(z\right)\right\}.$$

Using the log-link, the smallest extreme-value distribution becomes the exponential distribution when $\sigma = 1$, and it becomes the Rayleigh distribution when $\sigma = 0.5$. For all values of σ, the smallest extreme-value distribution with a log-link becomes the two-parameter Weibull distribution in which

$$f\left(z\right) = \frac{1}{\sigma\exp\left(x\beta\right)}\left(\frac{z}{\exp\left(x\beta\right)}\right)^{\frac{1}{\sigma}-1}\exp\left(-\left(\frac{z}{\exp\left(x\beta\right)}\right)^{\frac{1}{\sigma}}\right).$$

Comparing this with (5.7), we have $\beta = \dfrac{1}{\sigma}$ (shape parameter) and $\eta = \exp\left(x\beta\right)$ (scale parameter). Now σ is the scale factor in (6.4), which can be interpreted as being equivalent to *fit$scale*, from which we obtain

$$\beta = \frac{1}{fit\$scale}. \tag{6.5}$$

Similarly, β is the vector of coefficients in (6.4), which can be interpreted as being equivalent to *fit$coeff*, from which we derive

$$\eta = \exp\left(fit\$coeff\right), \tag{6.6}$$

since $x = 1$. As the output parameters of *survReg* were not explicitly mentioned in the reference, confirmation of our interpretation has been sought from other sources.

Venables and Ripley (1999) use the following parameterisation of the Weibull survival function (page 367):

$$S(t) = exp\left(-\left(\lambda.t\right)^{\alpha}\right),$$

where λ has the same usage as in the S-Plus manual and is the same as $1/\eta$ in our original parameterisation in (5.8), whilst α has been used in place of p in the S-Plus manual, this being equivalent to our β . (It can be appreciated that confusion with the Weibull distribution is partly due to the inconsistency in the function's parameterisation.) Venables and Ripley consider the hazard function in their section on parametric models for survival data (page 373–), and they discuss only models based on co-variates. An interpretation for a model without co-variates is required. They stated (page 375) that for the Weibull model, the scale factor estimate is $1/\alpha$. Interpreting 'scale factor' as *fit$scale*, we have $\alpha = 1/fit\$scale$, or in terms of our β, $\beta = 1/fit\$scale$, confirming (6.5). However, as this was not explicitly stated, and no interpretation was offered for the meaning of *fit$coeff*, we have sought verification from first principles.

Starting with (5.8), a linear model relationship can be established by transforming the scales.

$$\begin{aligned} S(t) &= \exp\left\{-\left(\frac{t}{\eta}\right)^{\beta}\right\} \\ \log\frac{1}{S(t)} &= \left(\frac{t}{\eta}\right)^{\beta} \\ \log\log\frac{1}{S(t)} &= \beta\left(\log t - \log\eta\right). \end{aligned}$$

Letting $x = \log\log\left(1/S(t)\right)$ and $y = \log t$, then

$$\begin{aligned} x &= \beta\left(y - \log\eta\right) \\ y &= \frac{x}{\beta} + \log\eta. \end{aligned} \tag{6.7}$$

Using S-Plus to assign a linear model of y in terms of x, we obtain two coefficients, which can be obtained using the S-Plus function *summary*. The interpretation of the output of this function is given on pages 169-71 in the S-Plus 2000 Guide to Statistics, Volume 1. The coefficient labelled "Intercept" is simply the intercept ($\log \eta$) of the linear (6.7), whilst the coefficient labelled with the variable name, x in our case, is the gradient $\left(\frac{1}{\beta}\right)$. Using a random sample of 1000 failure times generated from a Weibull distribution with $\beta = 3$ and $\eta = 100$, we have shown in Figure 6.2 the outputs from assigning $fitLm$ to the linear model, and fit to the model obtained using *survReg.* (See accompanying compact disc for the S-Plus script, *SurvRegParms.ssc*, used to generate the random sample and fit the linear and survival models.) Comparing these outputs, it can be seen that the Intercept in $fitLm$ is equivalent to the Intercept in fit. This component of the *survReg* model can be retrieved by *fit$coeff*. This leads to the relationship for the intercept

$$\begin{aligned} \log \eta &= \mathit{fit\$coeff} \\ \eta &= \exp\left(\mathit{fit\$coeff}\right), \end{aligned}$$

which confirms (6.6). It can be seen that the Log(scale) component of the *survReg* model (fit) is close in value to the logarithm of the x-coefficient of the linear model ($fitLm$), with the difference being attributable to the different methods being used in the functions to estimate these quantities. The component of the *survReg* model that extracts this component is *fit$scale*, which is equivalent to the exponential of the Log(scale) component of the *survReg* model. This leads to the relationship for the gradient

$$\frac{1}{\beta} = \mathit{fit\$scale},$$

and so

$$\beta = \frac{1}{\mathit{fit\$scale}},$$

which confirms (6.5).

Equations (6.6) and (6.5) can be verified numerically from the fact that the random sample has been generated using $\beta = 3$ and $\eta = 100$. This has been done in the S-Plus script *SurvRegParms.ssc* (see accompanying compact disc), the output of which is shown in Figure 6.2. The estimates of $\hat{\beta} = 2.980372$ and $\hat{\eta} = 98.86798$ are in close agreement to the assigned values. This supports the results obtained in (6.6) and (6.5).

An alternative S-Plus function is available for survival analysis, namely *censorReg*. Venables and Ripley (1999) discuss this function (page 379), and compare the parameter estimates obtained using *censorReg* with those obtained using *survReg*. They conclude that the estimates are the same. The S-Plus script *CensSurv.ssc*, which is included on

```
> summary(fitLm)

Call: lm(formula = y ~ x, data = xy, na.action = na.omit)
Residuals:
     Min        1Q  Median        3Q    Max
 -0.2582 -0.003091 0.00211 0.007531 0.1225

Coefficients:
                Value Std. Error   t value  Pr(>|t|)
(Intercept)    4.5942     0.0007 6934.5437    0.0000
          x    0.3435     0.0005  722.8242    0.0000

Residual standard error: 0.01911 on 998 degrees of freedom
Multiple R-Squared: 0.9981
F-statistic: 522500 on 1 and 998 degrees of freedom, the p-value is 0

Correlation of Coefficients:
  (Intercept)
x 0.41
> summary(fit)

Call:
survReg(formula = Surv(time, status, type = "right") ~ 1, data = randSample,
    na.action = na.exclude, dist = "weibull", scale = 0, control = list(maxiter
     = 30, rel.tolerance = 1e-005, failure = 1))
              Value Std. Error      z p
(Intercept)   4.59      0.0112  411.5 0
 Log(scale) -1.09       0.0249  -43.9 0

Scale= 0.336

Weibull distribution
Loglik(model)= -4889.7   Loglik(intercept only)= -4889.7
Number of Newton-Raphson Iterations: 6
n= 1000

Correlation of Coefficients:
             (Intercept)
Log(scale) -0.311

> 1/fit$scale
[1] 2.980372
> exp(fit$coeff)
 (Intercept)
    98.86798
```

Fig. 6.2. *Comparison of outputs from a linear model and a Weibull model.*

the accompanying compact disc, has been written to compare the parameter estimates of *fit$coeff* and *fit$scale* obtained from each of the two S-Plus functions using several components from our database. In the majority of cases there is close agreement in the respective values. However, for some components there is a significant difference. A difference as high as 54% has been found in the estimates of *fit$scale*, with a corresponding difference of 12% in *fit$coeff*. A correlation between the number of failures in the sample and the degree of agreement in the estimates has not been found. The difference between the estimates appears to be inexplicable. We have chosen to use *survReg* in our analysis as this was based on a function that has been tried and tested for a much longer time than the newer *censorReg* function. In addition, *survReg* is able to produce an output much more quickly than *censorReg*.

6.5.2 Point Estimate: Implementation in S-Plus

Having determined the meaning of the output of the *survReg* function, estimates of the warranty cost and its variance using Weibull modelling can now be obtained. A point estimate of the warranty cost, w_p, of component p can be calculated using (5.32). As for exponential modelling, an estimate of the cost of replacing the component is provided by the mean cost of repair, $\bar{c}_p$. In order to obtain an estimate of the expected number of replacements, the *survReg* function is firstly used to fit a Weibull model to the survival data using maximum likelihood. From the model parameters, the expected number of replacements is estimated using Xie's (1989) numerical methods. As discussed in Section 5.6.3, this involves the solving (5.25) and (5.24). In order to achieve this, the S-Plus function, *fRenew*, which is included in Appendix C.4, has been written.

The renewal program *fRenew* has been verified by comparing the results of this function with the output from Murthy and Iskandar's (1992) Fortran program. Their program's output contains four decimal places, which is identical to the results obtained using *fRenew*, correct to four decimal places.

Further verification has been sought by setting the Weibull shape and scale parameters to unity. This reduces the Weibull distribution to the exponential with a failure rate of one, so that $M(t) = t$. With this result, it is possible to test the accuracy of the *fRenew* function. It has been found that by setting a step size, h, to 0.01, it is possible to achieve an accuracy of five decimal places. Reducing the step size to 0.001 increases the accuracy to seven decimal places.

This study's data measures time in days for a warranty period of three years, which is equivalent to1096 days. A step size of 1 has been used in the analysis, which effectively breaks up the warranty period into 1096 intervals.

Further verification has been sought by comparing the results from *fRenew* with Baxter, Scheuer, Blischke and McConalogue's (1981) renewal tables. They used a cubic spline algorithm to produce their tables, and have claimed a discrepancy of no more than one digit in the fourth decimal place against the tables prepared by White (1964) and Soland (1968). Using the result that $M(t;\beta,\eta) = M\left(\frac{t}{\eta};\beta,1\right)$, Baxter et al. (1981) presented a comprehensive set of renewal tables with just a unity scale parameter. For a three-year warranty period, $t = 1096$ days. The data in this study has yielded estimates of the shape parameter, β, of the order of 0.9 to 3, and of the scale parameter, η, of the order of 4000 to 32000. Thus, the ratio of t/η was of the order of 0.03 to 0.3. Agreement to the four decimal places quoted in Baxter's et al. (1981) tables in this vicinity has been found, in most cases. The most notable deviations have been with very small values of t, around 0.05, when $\beta = 1.25$, and with small values of β (around 0.55), where agreement to three decimal places has been reached.

The function *fCstWbl* has been written to obtain the warranty cost estimates of (5.32), and is included in Appendix C.5. The function takes the failure data of one component at a time and uses the *survReg* function to fit a Weibull model to the survival data. Using the Weibull model parameters obtained by *survReg*, the expected number of replacements during the warranty period, $M_p(t_w)$, is then calculated using *fRenew*.

6.5.3 Variance: Implementation in S-Plus

When a Weibull model is fitted to the survival data using the S-Plus function *survReg*, a confidence interval on the two parameters of the distribution is returned. From these intervals, a confidence region for the number of renewals can be obtained using (5.13), which requires $var\left(\hat{\theta}\right)$, as given by (5.12). Assigning fit to the output of *survReg*, then $fit\$var[\cdot,\cdot]$ gives the variance-covariance matrix of *fit$coeff* and $\log$ (*fit$scale*), where the latter two components are related to the Weibull parameters as given by (6.5)) and (6.6).

Using the Taylor series expansion, we have

$$\begin{aligned}
g(x) &= g(\mu_x) + (x-\mu_x)\,g'(\mu_x) + \ldots \\
g(x) &\approx g(\mu_x) + (x-\mu_x)\,g'(\mu_x) \\
g(x) - g(\mu_x) &\approx (x-\mu_x)\,g'(\mu_x) \\
E\left[\{g(x)-g(\mu_x)\}^2\right] &\approx E\left[(x-\mu_x)^2\{g'(\mu_x)\}^2\right] \\
&\approx \{g'(\mu_x)\}^2\,var(x)\,.
\end{aligned}$$

That is

$$var\left[g\left(x\right)\right]\approx\left[g'\left(\mu_x\right)\right]^2 var\left(x\right). \tag{6.8}$$

To find the variance of η, let x refer to *fit$coeff*. Using (6.6), put $g\left(x\right) = \eta = \exp\left(x\right)$, so that $g'\left(x\right)=\exp\left(x\right)$. A value for μ_x is provided by *fit$coeff*. Thus

$$\begin{aligned} g'\left(\mu_x\right) &= \exp\left(\mu_x\right) \\ &= \exp\left(\mathit{fit\$coeff}\right) \end{aligned} \tag{6.9}$$

A value for $var\left(x\right)$ is obtained from $fit\$var\left[1,1\right]$, which is the variance of *fit$coeff*. Substituting into (6.8) an estimate of the variance of η can be given as

$$v\hat{a}r\left(\eta\right)=\exp\left(2\times \mathit{fit\$coeff}\right)\times fit\$var\left[1,1\right] \tag{6.10}$$

To find the variance of β, we use (6.5) can be used. However, $fit\$var\left[2,2\right]$ gives the variance of $\log\left(\mathit{fit\$scale}\right)$. Thus, we will put $y=\log\left(\mathit{fit\$scale}\right)$, rather than using y as *fit$scale*, before applying (6.8). Now

$$\begin{aligned} \beta &= \left(\mathit{fit\$scale}\right)^{-1} \\ &= \exp\left(-\log\left(\mathit{fit\$scale}\right)\right) \\ &= \exp\left(-y\right). \end{aligned}$$

Putting $h\left(y\right) = \exp\left(-y\right)$, then $h'\left(y\right) = -\exp\left(-y\right)$. Using $\mu_y = \log\left(\mathit{fit\$scale}\right)$, we have

$$\begin{aligned} h'\left(\mu_y\right) &= -\exp\left(-\mu_y\right) \\ &= -\exp\left(-\log\left(\mathit{fit\$scale}\right)\right) \\ &= -\left(\mathit{fit\$scale}\right)^{-1}. \end{aligned} \tag{6.11}$$

A value for $var\left(y\right)$ is obtained using $fit\$var\left[2,2\right]$. Substituting into (6.8) an estimate of the variance of β is given by

$$v\hat{a}r\left(\beta\right)=\left(\mathit{fit\$scale}\right)^{-2}\times fit\$var\left[2,2\right]. \tag{6.12}$$

Having obtained expressions for the variance of η and β, an expression for their covariance is needed. Now

$$\begin{aligned} cov(\eta,\beta) &= E\left[(\eta-\mu_\eta)(\beta-\mu_\beta)\right] \\ &= E\left[(g(x)-g(\mu_x))(h(y)-h(\mu_y))\right] \\ &\approx E\left[(x-\mu_x)g'(\mu_x)(y-\mu_y)h'(\mu_y)\right] \\ &\approx g'(\mu_x)h'(\mu_y)\times E\left[(x-\mu_x)(y-\mu_y)\right] \\ &\approx g'(\mu_x)h'(\mu_y)\times cov(x,y). \end{aligned}$$

Using (6.9) and (6.11), and $cov(x,y) = fit\$var[1,2]$, an estimate for $cov(\eta,\beta)$ can be given as

$$\begin{aligned} \hat{cov}(\eta,\beta) &= \exp(fit\$coeff)\times -(fit\$scale)^{-1}\times fit\$var[1,2] \\ &= -\frac{\exp(fit\$coeff)\times fit\$var[1,2]}{(fit\$scale)}. \end{aligned} \tag{6.13}$$

Equation (5.13) can now be implemented using (6.6), (6.5), (6.10), (6.12) and (6.13), to obtain a $100(1-\alpha)\%$ confidence region for the parameter set (η,β). Equation (5.13) can be written as

$$\begin{bmatrix}\eta-\hat{\eta} & \beta-\hat{\beta}\end{bmatrix}\begin{bmatrix} var(\eta) & cov(\eta,\beta) \\ cov(\eta,\beta) & var(\beta)\end{bmatrix}^{-1}\begin{bmatrix}\eta-\hat{\eta} \\ \beta-\hat{\beta}\end{bmatrix} \leqslant 2F(1-\alpha,2,n-3).$$

A set of points that lie on the boundary of this inequality can be found by using the output of fit. This is done by solving

$$\begin{aligned} &\begin{bmatrix}\eta-\exp(fit\$coeff) & \beta-(fit\$scale)^{-1}\end{bmatrix}\begin{bmatrix} fit\$var[1,1] & fit\$var[1,2] \\ fit\$var[1,2] & fit\$var[2,2]\end{bmatrix}^{-1} \\ &\qquad\times\begin{bmatrix}\eta-\exp(fit\$coeff) \\ \beta-(fit\$scale)^{-1}\end{bmatrix} = 2F(1-\alpha,2,n-3), \end{aligned}$$

which has been implemented in S-plus in the function *fConfReg*, presented in Appendix C.6. These points provide a $100(1-\alpha)\%$ confidence region for the parameter set (η,β). This can then be used in the renewal function *fRenew* (using the techniques discussed in Section 5.6.3,) to obtain a confidence interval for the expected number of failures during the warranty period.

The variance of the expected number of renewals can then be obtained using (5.26). For component p, $M_{pi}(t_w)$ can be evaluated for each point i on the confidence region by passing the point to the function *fRenew*. The point estimate of the expected number of renewals, $\hat{M}_p(t_w)$, is obtained by passing the estimated values of the Weibull parameters to *fRenew*, as discussed in Section 6.5.2. The number of points in the confidence region is given by n.

The variance of the cost of component replacement, $var\ (c_p)$, is given by the variance of the cost of replacing component p. The variance of the warranty cost can be obtained using (5.34). This is implemented in the S-Plus function *fCstWbl*, which is presented in Appendix C.5.)

As the function *fCstWbl* is processor intensive, taking about one hour per component on an AMD1800, the script *CallCostWbl.ssc* has been written to run the *fCstWbl* function by taking selected components at a time. The script is included on the accompanying compact disc.

6.5.4 Results

For a table of results from the individual-components analysis, see the data frame *costWbl* on the accompanying compact disc. The Weibull model has provided a good fit to most of the components with at least twenty claims, as discussed in Section 6.3.

For most components, the Weibull model has produced standard deviations of the warranty cost that are too large to be of any practical use. In only 59 of the 347 cases are the standard deviations less than the actual warranty costs themselves. This is due, in part, to the large variances in repair cost of some components. However, the exponential models are also affected by these large variances, and their total warranty cost variances are much smaller. In fact, the standard deviation of the number of renewals was larger than the number of renewals themselves in all but 69 of the 347 components. Thus, the large variances in component-repair costs could only account for ten more cases of standard deviations in component warranty cost being greater than the actual component warranty cost, above the number of cases due to large variances in the number of renewals.

In conclusion, the Weibull model does not provide a variance that could be used to evaluate confidence limits on the warranty costs. This points to a need to further develop a means of evaluating the variance under this model. In order to obtain workable confidence limits on the warranty costs in the current study, the variances obtained from the exponential models, as discussed in Section 6.4.3, and the variance obtained from the use of simulation, to be discussed in the next Chapter, will be used.

A discussion of the total warranty cost under the Weibull model, and a comparison of this cost with the one obtained under the two approaches using the exponential model, follows in the next section.

6.6 Comparison of Exponential and Weibull Models

To minimise processing time, and to facilitate the comparison of results, the two functions that generate the component warranty costs, namely *fCstExp* and *fCstWbl*, have been combined into the one function, *fCstExWb*. However, there is a difference in the processing in the functions *fCstWbl* and *fCstExWb*. As discussed in Section 6.3, there are fourteen cases where the Weibull model does not fit the data, so one or more of the oldest units have been trimmed off the data base to create a better fitting model. The function *fCstExWb* accommodates these changes, whilst the function *fCstWbl* does not. The combined function, *fCstExWb*, is included in Appendix C.7. This function is called from the script *RunCostExpWbl.ssc*, which can be found on the accompanying compact disc.

Only the exponential model has been fitted to components with less than twenty claims. As discussed previously, this has been done because it is felt that a reasonable number of observations are needed to fit a Weibull model. It is also justified in terms of the computer-intensive calculations needed to fit a Weibull model, bearing in mind that a large number of components exist with fewer than twenty claims. Both exponential and Weibull models have been fitted to the 347 individual components with at least twenty claims, so that the models can be directly compared. The results can be seen in the data frame *costExpWbl* on the accompanying compact disc. The data frame contains the log likelihoods, estimates of the warranty costs and the standard deviations for each component. The table reveals that the Weibull model provides a better fit than the exponential in all of the 347 cases, but that the differences in the log likelihoods are generally small. Thus, the exponential model also provides a suitable fit. There has been a problem with the Weibull model for most of these parts in that the standard deviations of the individual warranty costs are far too large to be useful. In contrast, the standard deviations from the exponential models are much more reasonable. Under the Weibull model, only in 59 of the 347 cases are the standard deviations less than the warranty costs themselves.

Table 6.5. *Comparison of exponential and Weibull warranty cost standard deviations.*

Part	PIPE AS.EX.FR.M	GSKT CYL HEAD
Claims	131	133
Exp. Cost	\$2.355	\$2.238
Exp. SD	\$0.516	\$1.118
Wbl Cost	\$2.355	\$2.234
Wbl SD	\$2.765	\$1.445
Exp loglikehd	-1760.5	-1785.3
Wbl loglikehd	-1752.7	-1731.7
β	0.72	2.87

Let us examine a couple of components to illustrate the situation. Table 6.5 shows the warranty costs and standard deviations of two components ("PIPE AS.EX.FR.M"–row 2979 in *costExpWbl* and "GSKT CYL HEAD"–row 2980 in *costExpWbl*), together with their log likelihoods. It can be seen that for both components, similar warranty costs are predicted by either model, but that the standard deviation from the Weibull model is larger than that of the exponential model. For the first component, the standard deviation from the Weibull model is larger than the cost, and does not provide a very workable figure. For the second component, the standard deviation from the Weibull model, although larger than the one from the exponential model, is a workable figure. Similarly, for all of the 59 cases where the standard deviations are less than the warranty costs under the Weibull model, the standard deviations under the exponential model are smaller. Table 6.5 also shows that the log likelihoods for each component are similar for the two different models.

Thus, one would be tempted to use the exponential model in preference to the Weibull. However, if the exponential model provided as good a fit as the Weibull, then one would expect that the Weibull parameter, β, of the fitted model would have a value near one. For the components discussed in the previous paragraph, the values of β are clearly not one. Therefore, one can conclude that the Weibull model provides a better fit than the exponential model for these two components. The Weibull model would certainly be preferable if warranty costs were to be extrapolated beyond the current warranty of three years.

A study of the values of the Weibull parameter β reveals that a range of values between 0.1457 and 33.08 exists for components with at least twenty claims. Five of the 348 components have β estimates that lie in the interval [0.99.1.01], indicating that the Weibull distribution is close to being exponential. This number increases to 25 when the interval is expanded to $[0.95, 1.05]$. A sensitivity analysis on the parameters of the models is discussed in the next chapter.

6.7 Model Estimates of Warranty Cost

6.7.1 Point Estimate

In this section, the warranty cost, variance and a 95% confidence interval are estimated using the three different models of failure discussed previously. The first model is based on individual components failing exponentially. The second also uses the exponential distribution, but all components with the same number of claims are taken to have the same failure rate. The third uses the Weibull distribution to model the failure of individual components with at least twenty claims, and the exponential distribution to model the failure of individual components with less than twenty claims. The total warranty cost, its variance and a 95% confidence interval have been calculated by the function *fWrntCstTot*, which is presented in Appendix C.8.

The warranty cost per vehicle, W, can be estimated using (5.36). Before applying (5.36), a method for handling the 4,715 cases where no part number is recorded in the database needs to be determined. These repairs are cases of adjustments where no part has been replaced. The nature of the adjustment has not been recorded, and so these repairs must be treated as one-off claims and modelled individually. As such, the failure times can be modelled exponentially, and the warranty cost can be given by (6.1). Since each repair is modelled separately, $r_p = 1$. An estimate of λ_p is given by λ_1, the estimated failure rate of one-claim components obtained in the function *fCstExpGp*. This is the failure rate when the single failure occurs in the middle of the warranty period. Putting w_b as the warranty cost per vehicle of these "blank" parts, $c_{b,j}$ as the cost of repairing the jth "blank" component, and r_1 as the number of "blank" repairs, we have

$$w_b = \lambda_1 t_w \sum_{j=1}^{r_1} c_{b,j},$$

The one-claim failure rate is given by $\lambda_1 = 3.02458 \times 10^{-8}$, the warranty period in days is $t_w = 1096$, and the number of "blank" repairs is $r_1 = 4715$. The sum of the cost of these "blank" repairs can be obtained using the component cost field (*CmpCst*) in the *costExpWbl* data frame. However, as this value is the mean cost of all 4,715 repairs, it must be multiplied by the number of repairs to obtain the sum of the cost of all these repairs. Thus,

$$\begin{aligned} w_b &= 3.02458 \times 10^{-8} \times 1096 \times \$112.535 \times 4715 \\ &= \$17.59, \end{aligned}$$

which represents the warranty cost per vehicle of all the "blank" repairs.

Under the first model, where failure of individual components is modelled exponentially, the total warranty cost is the sum of the warranty cost of the individual components. This can be obtained by summing the *WrntExp* field in the *costExpWbl* data frame, excluding the record with no part recorded. These repairs have been treated separately, as discussed in the preceding paragraph, and then have been added to the total. The total warranty cost with this model is $272.84 per vehicle for the three-year warranty.

Under the second model, where components with the same number of claims are assigned the same failure rate, the total warranty cost is the sum of the warranty costs of all the groups. This can be obtained by summing the *Wrnt* field in the *costExpGp* data frame, which does not contain a record for "blank" components. The total warranty cost is obtained by adding the warranty cost for the grouped components together with the warranty cost of the "blank" components. This total is $266.91 per vehicle for the three-year warranty.

Under the Weibull model, the failure of components with less than twenty claims has been modelled exponentially. This has been done in the same way as the first model. Failure of components with at least twenty claims has been modelled by the Weibull distribution, and "blank" components have again been modelled separately. The total warranty cost under this model is obtained by adding the warranty cost for the Weibull model with the warranty cost of the "blank" components. The total cost under this model is $272.62 per vehicle for the three-year warranty.

The actual total warranty cost that has been recorded on the manufacturer's database is $8,302,477 for the three-year warranty. With a production of 30,138 in 1997, the warranty cost per vehicle is $275.48. This figure is in close agreement with the figures obtained from all three models. Thus, it can be concluded that all three models provide a reasonable point estimate of the current warranty, and hence, the additional process-

Table 6.6. *Total warranty cost, variance and confidence limits.*

	Warranty Cost	Standard Deviation	95% C.I.
Exponential	\$272.84	\$9.8278	[\$253.58, \$292.11]
Grouped Exp	\$266.91	\$9.2278	[\$248.82, \$285.00]
Weibull	\$272.62	\$5185.17	[\$0, \$10435]

ing to evaluate the Weibull model does not seem justified to model the cost of the current three-year warranty.

6.7.2 Variance

The variance of total warranty cost, $var\,(W)$, is given by (5.37), namely

$$var\,(W) = \sum_p var\,(w_p)\,.$$

As for the actual warranty cost, the variance of the total warranty cost is obtained by summing the individual variances. This has been done for each of the three models in the same way as it was done for the cost, with "blank" components being treated separately. The results of the calculations can be seen in Table 6.6.

Lower and upper 95% confidence limits in Table 6.6 have been obtained using critical values from the normal distribution with density

$$f\,(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-\,(t-\mu)^2}{2\sigma}\right],$$

where μ is the mean of the distribution and σ^2 is the variance. Thus, a 95% confidence interval for the total warranty cost, W, is given by

$$W \pm 1.9600S,$$

where S is the standard deviation of the cost, that is, the square root of the sum of the variances.

Confidence intervals using the gamma distribution with density

$$f\,(t) = \frac{\lambda^\beta t^{\beta-1}}{\Gamma\,(\beta)} \exp\,(-\lambda t)\,,$$

where λ is the rate parameter and β is the shape parameter, have also been obtained. Since $E\,(T) = \beta/\lambda$ and $var\,(T) = \beta/\lambda^2$, we have

$$\lambda = \frac{E\,(T)}{var\,(T)}$$

and

$$\beta = \frac{E^2(T)}{var(T)}.$$

A 95% confidence interval can then be obtained by finding the $[0.025, 0.975]$ quantile interval for the gamma distribution with the above parameters. Under the exponential model, the confidence interval is $[253.58, 292.11]$. For the grouped exponential model it is $[248.82, 285.00]$. These two confidence intervals are very close to the first two shown in Table 6.6, which are obtained using the normal distribution. The Weibull model does not provide suitable gamma distribution limits.

The chi-square distribution was also considered when estimating a 95% confidence interval. However, as the chi-square distribution is a special case of the gamma distribution with $\lambda = \frac{1}{2}$ and $\beta = \frac{v}{2}$, where v is an integer, it is not necessary to pursue this distribution further, as it is encompassed within the gamma distribution.

We conclude that the most suitable estimate of the total warranty cost is obtained by the Weibull model ($272.62 per vehicle) and that a 95% confidence interval for this value can be obtained using the individual exponential model's standard deviation and normal limits. The result is the interval $[253.58, 292.11]$.

6.8 Costing an Extended Warranty

In the preceding sections, three models were discussed, namely, the individual-component exponential model, where the failure of individual components was modelled using the exponential distribution, the grouped exponential model, where the failure of components with the same number of claims was modelled together using the exponential distribution, and the Weibull model, where the failure of individual components was modelled using the Weibull distribution. They were used to model the cost of the current warranty of three years. In this section, these three models are used to estimate the cost of the warranty should it be extended to five years.

The distance limitation will not be considered in this estimation. If it were to remain at 100,000 kilometres, it would cover the average user of 20,000 kilometres/year. (See discussion in Section 5.2.8.) However, in this case the warranty costs predicted by the models should be over estimates since some claims would be eliminated by the distance limitations of the warranty before the time limitation was reached. Alternatively, the

Table 6.7. *Total cost, variance and confidence limits of a 5-year warranty.*

	Warranty Cost	Standard Deviation	95% C.I.
Exponential	\$454.57	\$16.37	[\$422.48, \$486.67]
Grouped Exp	\$444.69	\$15.37	[\$414.56, \$474.82]
Weibull	\$770.86	$\$3.9826 \times 10^{12}$	—
Adjusted Weibull	\$589.59	$\$3.9826 \times 10^{12}$	—

distance limitation could be increased proportionately to 166,667 kilometres, a somewhat awkward figure to work with. This could be rounded down to 150,000 kilometres, a more workable figure, but keeping the distance limitation to the current 100,000 kilometres would seem to be the fairest option for the average user. By so doing, the lighter user would not subsidise the heavier user to the extent they are under the current warranty, and the manufacturer would have 'safer' cost predictions to work with.

Extrapolating the warranty costs to five years is a matter of using the models with a new time limit. As the model parameters have already been estimated in the previous sections, they can be used with the new time limitation in the models. This has been implemented in the function *fExtWrnt* (see Appendix C.9). The calling script for this function, *CallExtdWrnt.ssc*, is on the accompanying disc.

The results of this function can be seen in the data frame *ExtWrnt*, which is included on the accompanying compact disc. It can be seen that the warranty costs of components modelled by the exponential distribution have increased linearly, since the exponential model uses a constant failure rate. However, components that have been modelled by the Weibull distribution have increased in varying ways, depending on the values of the Weibull parameters. Similarly, the changes in the standard deviations depend on the values of the parameters.

The total warranty costs using the three models can again be obtained by the function *fWrntCstTot*. The figures are shown in the first three rows of Table 6.7. It can be seen that both the warranty cost and the standard deviation for the two exponential models are a linear extrapolation of the three-year warranty figures shown in Table 6.6. This, of course, is because the exponential models use a constant failure rate. The warranty cost under the Weibull model, however, has increased at more than a linear rate.

The ratio of extended warranty cost to current warranty costs of individual components has been examined. These ratios are unrealistically large for some components. In the most extreme case, the ratio is more than one thousand. Components with large ratios are characterised by large values of β, (the maximum being 16.70,) and relatively small values of η, compared with the other components. Thus, these components' failures are concentrated within a narrow band, resulting in a steep increase in the cumulative number of failures over a short period of time. When the majority of claims occur near the end of the warranty period, such as may be the case when non-urgent claims are held over to the end of the warranty, a steep rise in the failure rate can occur at that time. In such cases, the timing of the actual failure does not coincide with the occurrence of the claim, as has been assumed in this study. Thus, the modelling of component reliability will be inaccurate, and consequently, so will the modelling of the warranty cost. When the ratio is more than five, the Weibull model has been deemed to be inappropriate, and the exponential model has been used instead. This occurs in thirty-nine of the 347 components that have been modelled by the Weibull distribution. These thirty-nine components have the highest values of β.

The results of the adjusted Weibull model are shown in the fourth row of Table 6.7. It can be seen that this change lowers the total warranty cost to a more appropriate level. The standard deviation of the cost has not been affected. The acceptable level for the ratio of the extended warranty cost to the current warranty cost has been somewhat arbitrarily set to five. If this ratio is reduced to less than five, then clearly the warranty cost under the Weibull model would be further reduced, as more components would be modelled by the exponential distribution. An alternative approach to setting this ratio is to set an acceptable level of β before the Weibull model is rejected in preference to the exponential model. The corresponding value of β when the ratio is set to five is 3.142. Thus, any components that have a β-value larger than 3.142 have been modelled by the exponential distribution.

We conclude that the adjusted Weibull model provides the best point estimate of the total warranty cost. The model estimate is \$589.59. A 95% confidence interval cannot be obtained from the variance estimate of this model. Nor can the confidence interval obtained by using the exponential model be used, as it does not encompass the point estimate obtained by the Weibull model. However, a crude confidence interval can be procured using the standard deviation of the exponential model and a normal approximation to the distribution of point estimates. That is, a 95% confidence interval on the Weibull model can be given by

$$\mu_W \pm z_{0.975} sd_e,$$

where μ_W is the point estimate of the Weibull model ($589.59) and sd_e is the standard deviation obtained by the exponential model, the value of which is $16.35. Thus, the confidence interval is $[557.55, 621.63]$.

6.9 Sensitivity Analysis

In this section, a sensitivity analysis on the estimated Weibull parameters for single components is conducted. The change in warranty cost when the Weibull shape parameter, β, and the scale parameter, η, vary from their estimated values individually and jointly is examined.

6.9.1 Range of Parameter Values

The range of β-values that have resulted from the Weibull modelling discussed in this chapter is 0.1457 to 16.70, with the lower and upper quartiles being 1.113 and 2.283 respectively. The values of η range from 1,695 to 3.495e+023 (a somewhat unrealistic figure), with the lower and upper quartiles being 12,720 and 321,400 respectively. These figures represent very reliable components. For example, the component "4SPK RAD/CASS" (row 2965 in *costExpWbl*), with a value of η of 325 878, a value close to the upper quartile, has a mean life of almost 900 years! Although this may not necessarily be the true mean value of the component, it does indicate that the component should last the life of the "average" vehicle. The component "BODY STEER/RACK" (row 2929 in *costExpWbl*), with an η-value of 12,804, which is near the lower quartile, has a mean life of 31 years. This again represents a very reliable component.

To test the sensitivity of the total warranty cost on the values of Weibull parameters, five parts representing the minimum, maximum, median and lower and upper quartiles of β-values have been analysed. The function *fSenstv* has been written to do the analysis, and can be found in Appendix C.10. The parts and their respective values of β and η are:

- "WHEEL ALIGNMENT", with $\beta = 0.145\,726$ (the minimum value of β) and $\eta = 3.495\,15 \times 10^{23}$ (the maximum value of η, row 2805 in *costExpWbl*);
- "BRG,FNT SPRING", with $\beta = 1.117\,3$ (approximately the lower quartile of β) and $\eta = 606\,180$ (row 2768 in *costExpWbl*);

- "ABSORB.ASS.R,R/", with $\beta = 1.601\,1$ (approximately the median value of β) and $\eta = 25\,443$ (row 3005 in *costExpWbl*);
- "GASKET KIT, POW", with $\beta = 2.2949$ (approximately the upper quartile of β) and $\eta = 16482$ (row 2921 in *costExpWbl*);
- "CHECK FR DR 083", with $\beta = 16.704$ (the maximum value of β) and $\eta = 1695.0$ (row 2713 in *costExpWbl*).

6.9.2 Results

The change in warranty cost as β changes can be seen by examining Figure 6.3. It can be seen that a decrease in the value of β produces a non-linear increase in the warranty cost. A 10% decrease in β results in an increase in warranty costs in the representative sample of between 65% and 107%. Conversely, a 10% increase in the value of β produces a decrease in warranty cost of between 46% and 52%. Thus, it can be seen that an underestimate of the shape parameter, β, will result in an overestimate in the warranty cost. Conversely, an overestimate in the value of β will result, to a lesser extent, in an underestimate in warranty cost.

Similarly, examination of Figure 6.4 reveals an inverse relationship between changes in warranty cost and changes in η. It appears to be almost linear for most components, but not all. A 10% decrease in η results in an increase in the warranty cost of components in the representative sample of between 1.5% and 480%. Conversely, a 10% increase in the value of η produces a decrease in warranty cost of between a small 1.4% and larger 80%. The figure also shows that the extremes in these variations are displayed by one component. Thus, it can be seen that an underestimate of the scale parameter, η, may or may not result in a significant overestimate in the warranty cost. Conversely, an overestimate in the value of η may or may not result in a significant underestimate in warranty cost, but the effect will be less than if η is underestimated.

Figure 6.5 shows the combination effect on warranty cost when both β and η are varied by $\pm 10\%$. The largest increases in warranty cost occur when both β and η are decreased by 10%, whilst the largest decrease in warranty cost occurs when both β and η

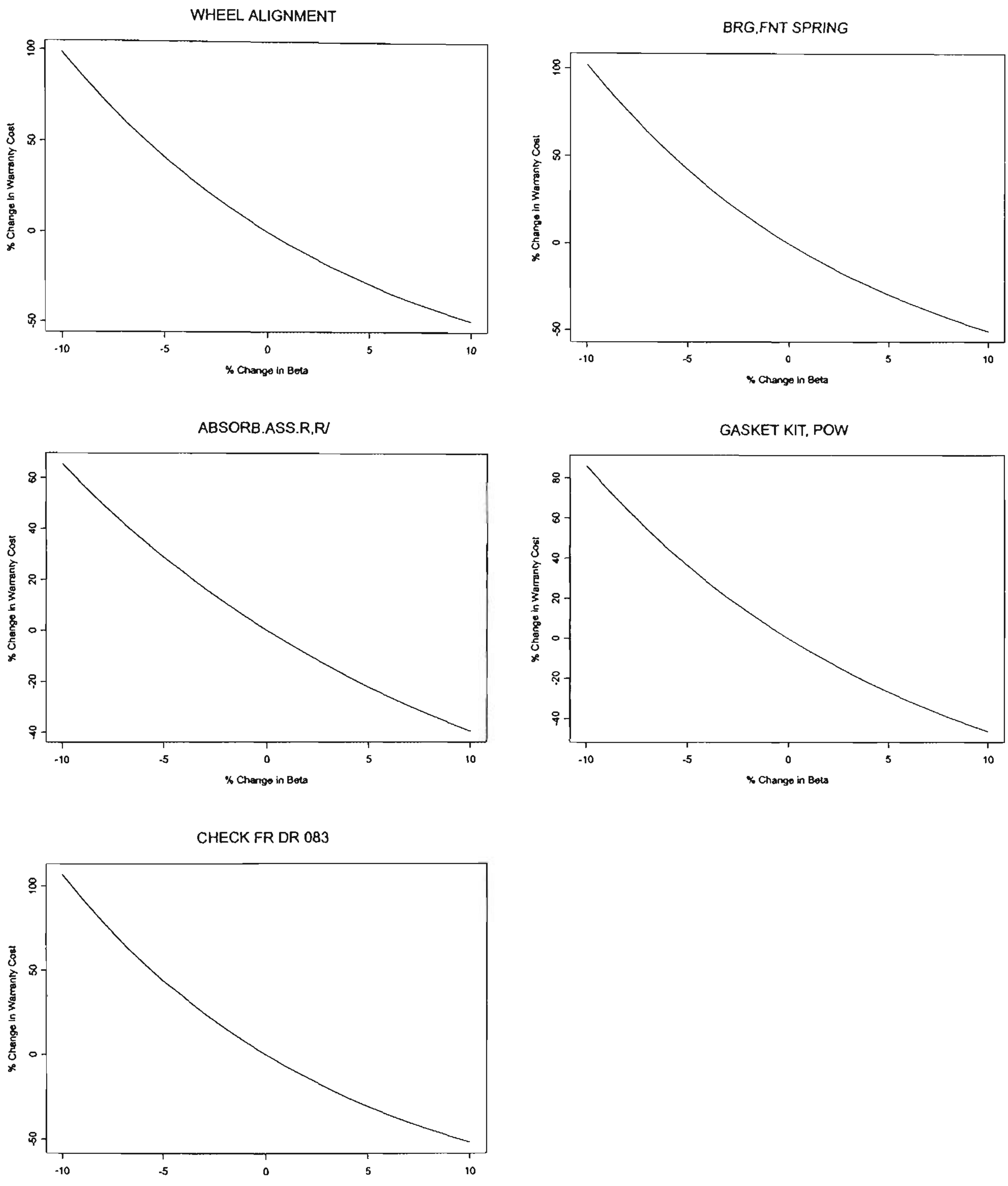


Fig. 6.3. *Percentage change in warranty cost against percentage change in* β.

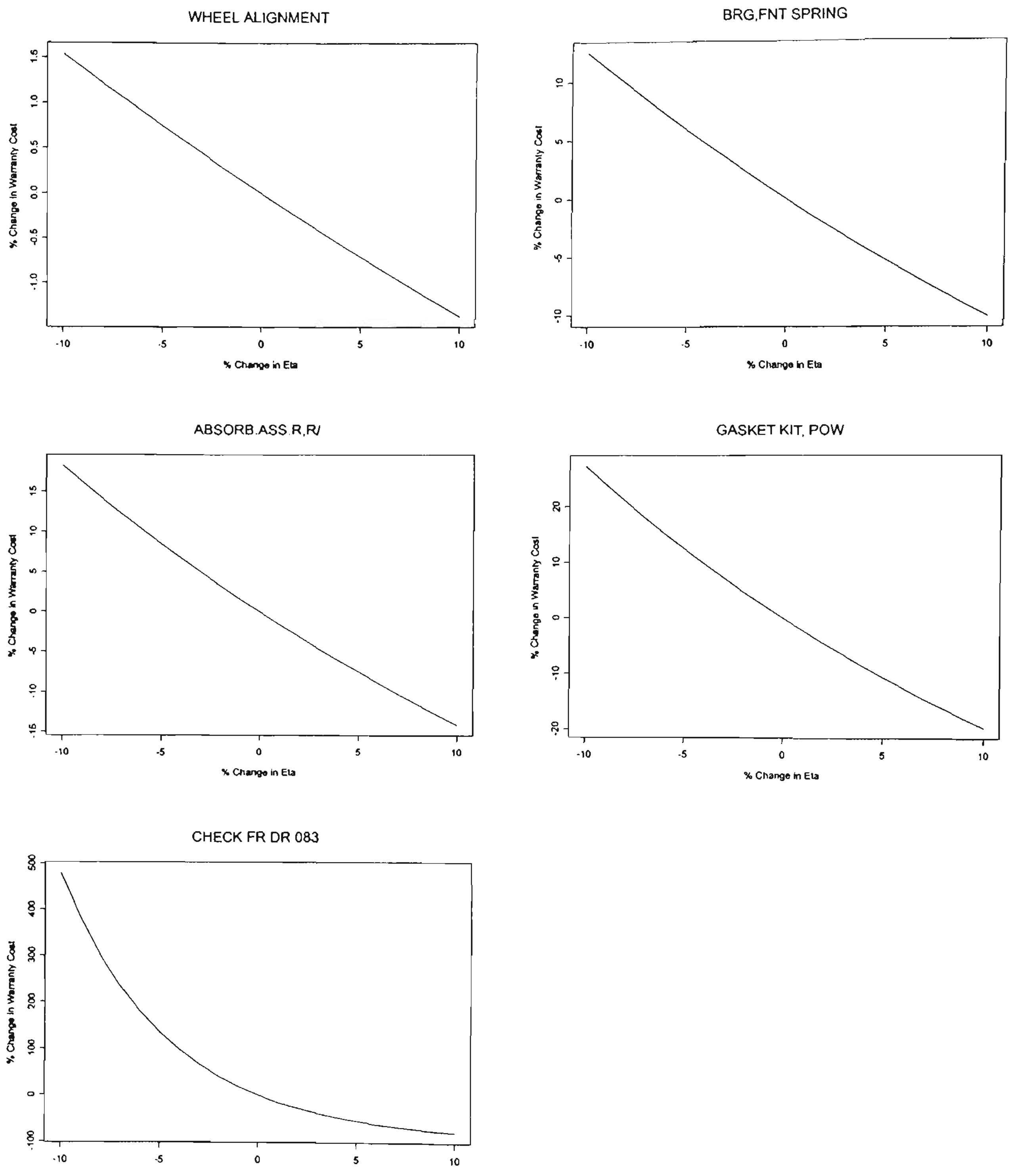


Fig. 6.4. *Percentage change in warranty cost against percentage change in* η.

are increased by 10%. Of the representative sample of components, the largest percentage change is seen in the last plot of Figure 6.5.

6.9.3 Interpretation of Results

The degree of sensitivity of the estimates of the Weibull parameters on the estimate of the warranty cost depends upon which component is being analysed. That is, for some components, there is little change in the warranty cost when the parameters vary, whilst for other components, there is a significant difference. Hence, it is important to look at each component individually in order to draw any conclusions about the sensitivity of the parameter estimates.

An overall trend, however, is that an underestimate in the value of the shape parameter, β, will result in an overestimate in the warranty cost of a component. Similarly, an underestimate in the value of the scale parameter, η, will also result in an overestimate in the warranty cost of a component. The converse of both of the last two statements is also true. For most, but not all, the estimate of the value of β is more sensitive than the estimate of the value of η.

6.10 Conclusion

This chapter has dealt with the implementation of the cost models developed in the previous chapter, using the manufacturer's warranty data. Issues, such as how to deal with components having only a few records of claims, or how do deal with records with no part identification, have been discussed. Three different ways of modelling the data have been presented: the individual-component exponential model, where the failure of individual components has been modelled using the exponential distribution; the grouped exponential model, where the failure of components with the same number of claims has been modelled together as a group, using the exponential distribution; and the Weibull model, where the failure of individual components has been modelled using the Weibull distribution. The cost per vehicle of the three-year warranty has been calculated under the three models, which have been found to be in close agreement. The variances obtained by the two exponential models are also in close agreement, but the variance obtained by the Weibull model is too large to be of any practical use for most components. The large

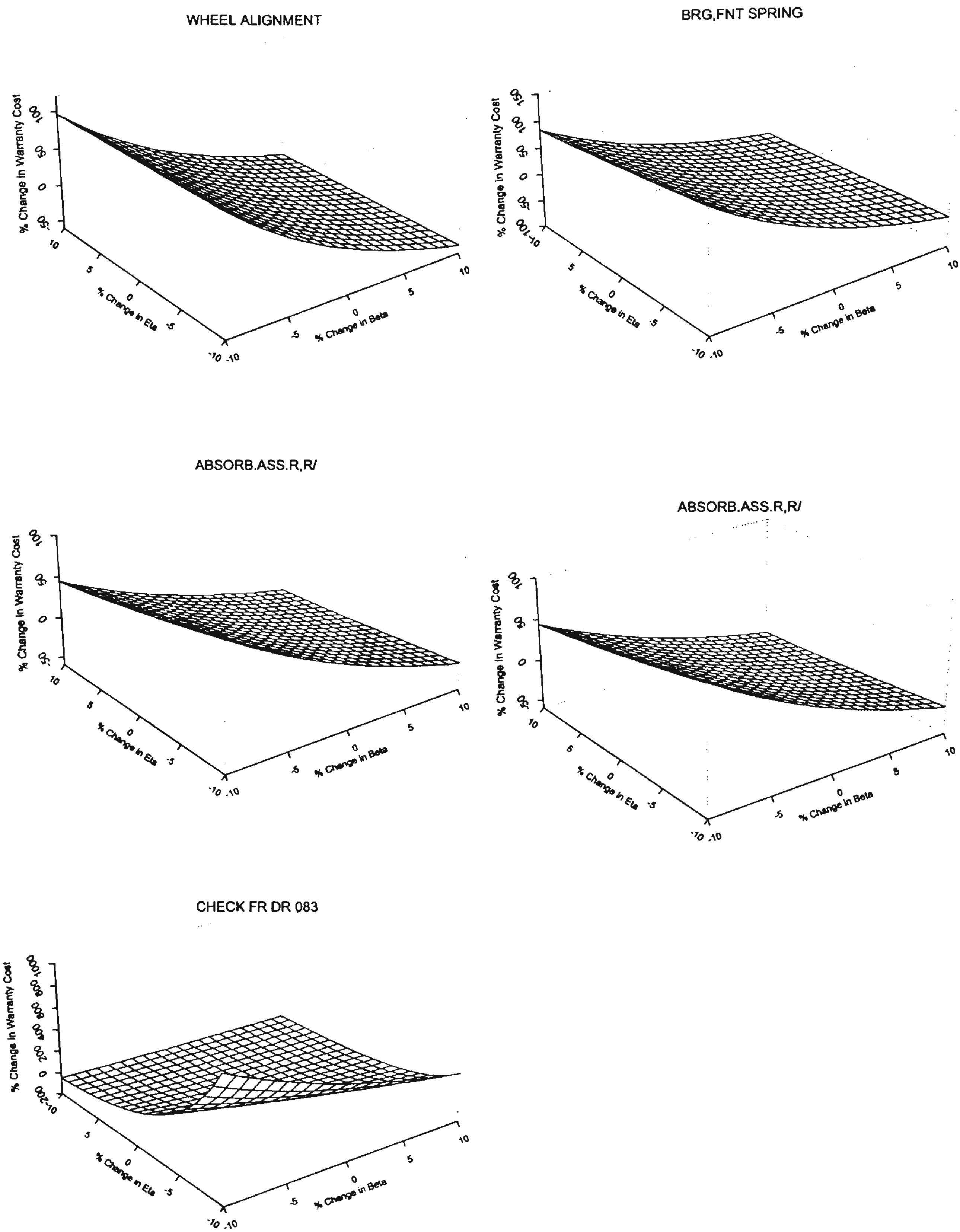


Fig. 6.5. *Percentage change in warranty cost against percentage change in β and η.*

variances obtained for the expected number of renewals for individual components under the Weibull model is the reason for the large variance in the overall Weibull model. Confidence intervals from each of the models have been presented, and further confidence intervals are obtained through the simulations discussed in the next chapter. The cost of extending the warranty to five years has also been discussed, as well as the variance and confidence intervals for each of the three models. The conditions under which the Weibull model would be inappropriate when extrapolating the cost models to five years have also been specified.

The sensitivity-analysis has shown that the warranty cost is more sensitive to changes in the shape parameter (β) of the Weibull model than it is to changes in the scale parameter (η). That is, a percentage change in the value of β produces a greater change in the warranty cost than does the same percentage change in the value of η. Thus, it seems to be more important to accurately estimate the value of β than it is to accurately estimate the value of η. However, the sensitivity of either parameter varies with each part, so it is important to look at each part individually before reaching conclusions about the sensitivity of each of the parameters.

The next chapter describes a number of simulations that are used to validate the models developed in this chapter.

Chapter 7
Simulation Studies

7.1 Introduction

In the last chapter, the manufacturer's warranty data were used to estimate the distribution of each components' reliability, its cost of repair, and subsequently, its warranty cost. Using a point estimate and the variance of each component's warranty costs, the total cost of the warranty was estimated, and a 95% confidence interval was obtained. This was done for both the three-year warranty and an extended five-year warranty. However, the Weibull model did not provide a usefull confidence interval, as the variances of the component costs were too large as a result of the large variance in the number of renewals.

In this chapter, simulations are used to obtain confidence intervals for the warranty cost per vehicle using the Weibull models of component reliability. Simulated failures times are generated over a warranty period using the parametric models. By repeating the simulations hundreds of times, it is possible to obtain point and interval estimates for the cost of the warranty per vehicle. From the simulated data, it is possible to see whether the parametric models produce results that are close to the observed results.

In the first simulation, non-parametric bootstrap sampling is used to obtain the total warranty cost per vehicle for the three-year warranty. Thus, the model parameters are not used in this particular simulation. The second simulation uses parametric models to generate warranty costs per vehicle for both the three-year and five-year warranties. The third simulation also uses parametric models to generate data, as in the second simulation, except that a distance limitation is imposed upon the warranty in addition to the time limiation. This last simulation tests the validity of our warranty cost model, which has been based on time alone.

Apart from the above simulations which cover all the components in the database, this chapter also includes simulations of single-component failures. There are three simulations that test for the length of time needed for data to be collected before the parameters of a Weibull model can be accurately estimated. In the first of these simulations, all the products are the same age. In the second, continuous production is simulated, so new data is added to the database. In the third, continuous production is again simulated, but the database is restricted to the first year's production. These last three simulations have been carried out to test the validity of the methods used in the modelling in the previous chapter.

As in the previous chapters, a number of original S-Plus scripts and functions have been written by the author to do the simulations. Many of these are included in Appendix E. All the functions, scripts and data frames that are referred to in the text can be viewed on the accompanying compact disc.

7.2 Bootstrap Estimate of Total Warranty Cost

7.2.1 Purpose

The aim of the simulation in this section is to use the bootstrap method (see Section 5.8.4) to obtain a point and interval estimate of the warranty cost per vehicle from one year's production. The purpose of the simulation is to compare the results obtained from the parametric models of the last chapter, and to obtain a non-parametric confidence interval.

7.2.2 Method

Using the manufacturer's data, a database of vehicles produced during one year, and the warranty cost paid for each of these vehicles, is established. Then, bootstrap samples, the same size as the year's production, are randomly selected, with replacement. The statistic of interest from each of these bootstrap samples is the warranty cost per vehicle. This is simply the total warranty cost for the sample divided by the number of vehicles produced. Repeating the sampling 10,000 times provides a distribution of warranty costs per vehicle that is normally distributed. The mean of all the samples provides a point estimate of the warranty cost and the 2.5 to 97.5 percentile range can be used for a 95% confidence interval.

The script *BootCost.ssc* contains the S-Plus code to obtain the bootstrap estimates. The script is on the accompanying compact disc. In this script, the warranty cost of each of the 30,138 vehicles produced during 1997 is calculated and stored in the data frame *VINCost*. A random sample of 30,138 vehicles, representing one year's production, is drawn with replacement from *VINCost*. The total warranty cost of this sample is calculated and an average cost per vehicle is obtained. The sampling is repeated 10,000 times. The BC_a 95% confidence interval on the warranty cost per vehicle is obtained using the 2.5 and 97.5 percentiles of the 10,000 repetitions. (See Section 5.8.4 for a discussion of the BC_a confidence interval.)

7.2.3 Results

Both the observed and the bootstrap estimate of the warranty cost per vehicle are $275.48. The BC_a 95% confidence interval is [$269.07, $282.17]. It can be seen that the bootstrap estimate of the total warranty cost is in close agreement with the observed cost of $274.98, and with the costs predicted by the three models reported in Table 6.6 of Section 6.7. The bootstrap 95% confidence interval gives tighter bounds than those obtained by the two exponential models (disregarding the Weibull model bounds). This is because the models developed in the last chapter are based on the warranty data. The reliability of some components, and their cost of repair, exhibit skewness and large variances, which are characterised in the models of the last chapter. However, they will not be as evident in the bootstrap confidence interval because the costs per vehicle of the bootstrap samples are taken from the one fixed set of vehicles. The warranty cost per vehicle of the bootstrap samples will be normally distributed because of the large number of samples (10,000). This is clearly seen in Figure 7.1, and is confirmed by the QQ plot in Figure 7.2.

7.3 Simulated Warranty Costs: 3 and 5 Year Warranties

7.3.1 Purpose

The purpose of this simulation is to obtain point and interval estimates of the warranty cost per vehicle from simulated data, using the parameters determined in the last chapter. By so doing, it will be possible to establish whether the parameter estimates obtained in the last chapter result in similar point and interval estimates using the simulated data. It will

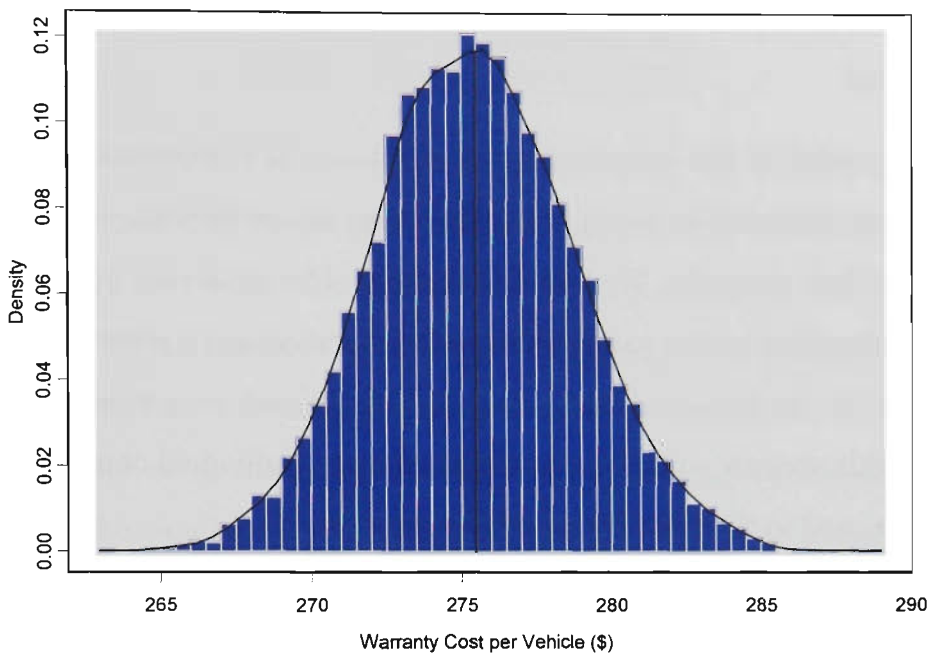


Fig. 7.1. *Distribution of warranty cost per vehicle of bootstrap samples.*

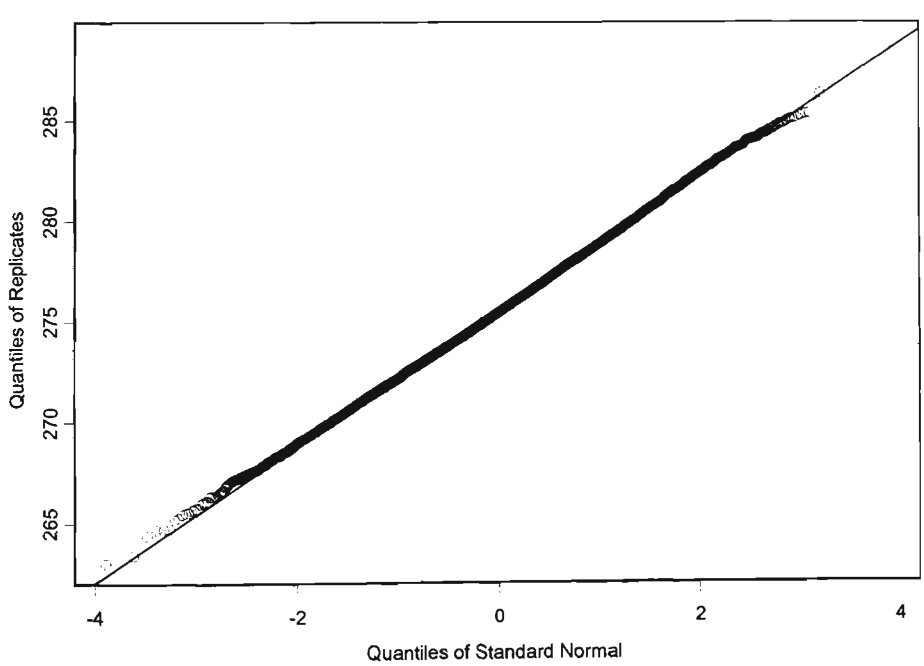


Fig. 7.2. *Plot of simulated sample quantiles against standard normal quantiles (QQ plot).*

also be possible to obtain an interval estimate for the Weibull model using the simulated results.

7.3.2 Method

In the Weibull model of the last chapter, the reliability of components with fewer than twenty claims are modelled exponentially, whilst that of components with at least twenty claims are modelled using the Weibull distribution. The reliability of components that have not been identified with a part number in the manufacturer's database are also modelled exponentially, as discussed in Section 1. This procedure is again followed in the simulations of this section, so that the model used on an individual component in the last chapter is again used in the simulations of this chapter.

Failure times for each component in each vehicle in one production year are simulated using the parameters that were obtained in the last chapter for that component, and are stored in the data frame *costExpWbl* (see Section 6.6). If a component fails within the warranty period, a new failure time is generated for the replacement component. This process continues until the warranty runs out for each vehicle.

The cost of repairing each failed component is obtained by randomly selecting, with replacement, the cost of repairing that component. These costs are obtained from the data frame, *compAge*, of component failure times and costs (see Section 4.5).

The warranty cost per vehicle in the production year is then calculated from the total cost of replacing the parts that failed during the warranty period. As this simulation is processor intensive, the number of simulations has been set to 500, which is sufficient to use the central limit theorem to make inferences about the mean value of the warranty cost per vehicle, and to obtain a confidence interval. The 95% confidence interval is made up of the 2.5 to 97.5 percentile range. This procedure is followed for both the three-year and five-year warranties.

The S-Plus function *fWCstSim* has been written to produce this simulation. Its script (*WrntCstSimul.ssc*) is included in Appendix E.1. The function is used to simulate the warranty cost for both the three-year and five-year warranties. The inputs of the function are the data frame of three-year warranty costs (*costExpWbl*), the data frame of five-year warranty costs (*extdWrnt*), the data frame of component fail times and costs (*compAge*), the length of the warranty, the number of vehicles produced in a year, the minimum number

Table 7.1. *Simulated warranty cost per vehicle.*

	Mean	95% C.I.
3-year	\$266.65	[262.27, 271.46]
5-year	\$580.34	[574.03, 586.38]

of claims a component must have for it to be modelled by the Weibull distribution, the threshold for the number of times larger the five-year warranty cost has to be above the three-year warranty cost before the Weibull model is rejected in favour of the exponential (see Section 6.8), and the form of the output that is required. The output can be in the form of either a data frame containing part number, the warranty cost and the number of replacements of each component, or it can simply be the overall warranty cost per vehicle that results from the simulation (the default). Since the creation of 500 repetitions takes a lot of computer processing, a script, *callWrntCstSimul.ssc* (see accompanying compact disc), has been written to call the function *fWCstSim*, place the results in the vectors *wrCstSim* and *wrCstSim5*, and to calculate appropriate statistics.

7.3.3 Results

Using the Weibull model for both the three-year and five-year warranties, the mean warranty cost per vehicle of the 500 simulations, and the 95% confidence interval, made up of the 2.5 to 97.5 percentiles, are shown in Table 7.1.

The cost per vehicle for the three-year warranty estimated by the Weibull model (Table 6.6) is \$272.62, and the actual warranty cost per vehicle for 1997 is \$275.48. Both figures are similar to the three-year estimate of \$266.65 obtained in the simulation. Likewise, the warranty cost per vehicle for the five-year warranty obtained by the adjusted Weibull model (Table 6.7) is \$589.59, which is also similar to the corresponding estimate of \$580.34 in Table 7.1. However, in both instances, the Weibull model estimates and the actual costs are outside the confidence interval obtained through the simulation. This could be explained by the skewness exhibited in the real data that is not displayed to the same extent in the simulation. Since the value of the parameters have been set to their point estimates in the simulations, the variation in the reliability of a component is as a result of the random variation around the set parameter. However, there are confidence intervals associated with each parameter estimate. If the parameter values in the simulation were to be determined by sampling within these confidence intervals, a greater variation in the warranty cost estimate would result. Therefore, the confidence intervals would be wider. Although this method has not been pursued in this study because of the time constraints, it is intended that this be pursued in a later study. For the purposes of this study, however, there is sufficient similarity between the model estimate of the last chapter and the simulated results, considering the simplifications made in the modelling. Therefore it can be concluded that the simulations support the results obtained by the Weibull model.

7.4 Simulated Warranty Costs: Age and Distance Limitations

7.4.1 Purpose

The purpose of this simulation is to extend the one of the last section by including distance, in addition to time, to the failure event. By so doing, the distance limitations of the warranty can also be imposed on the simulated data and a comparison of the warranty costs from the two simulations can then be made.

7.4.2 Method

The S-Plus function *fWCstSmD*, the script of which (*WrntCstSimulDist.ssc*) is included in Appendix E.2, has been written to run this simulation. Its structure is similar to *fWCstSim*, except that a distance to failure has been included in the simulated data. If the failure occurs after the distance component of the warranty expires, then a warranty claim is not made in the simulation. As in the previous simulation, the function *fWCstSim* is called using the script *callWrntCstSimul.ssc* (see accompanying disc), which houses the results in the vectors *wrCstSimD* and *wrCstSimD5*, and calculates appropriate statistics.

As in the last simulation, the number of repetitions has been set to 500 for the same reasons.

7.4.3 Results

Table 7.2 shows the results of the simulation. It can be seen that the mean warranty cost per vehicle with a distance limitation, for both the three and five-year warranties, are almost the same as the corresponding values for the mean warranty cost without the

Table 7.2. *Simulated cost per vehicle with distance limitations on the warranty.*

	Mean	95% CI
3-year	\$266.72	[262.52, 271.38]
5-year	\$580.41	[574.95, 587.02]

distance limitation, shown in Table 7.1. Thus, there has been little impact on the estimated warranty cost per vehicle by not using the distance limitation in building up the models in the last chapter. The reason this occurs was discussed in Section 5.2.8.

7.5 Reliability Estimation Using Same-Age Data

7.5.1 Purpose and Background

The purpose of the simulation in this section is to determine how old a product has to be before one can make an accurate estimate of the lifetime distribution of a component.

A similar simulation to the one in this section was reported by Summit, Cerone and Diamond (2002a).

The simulations in this section generate data for one year's production of 30,138 vehicles. The entire year's production is then allowed to age from one to five-years, in yearly increments. How accurately the reliability of a component can be estimated in each of these circumstances is tested.

The simulations reflect situations where a manufacturer:

- provides a one-year warranty, such as was offered by Australian car manufacturers in the early nineties;
- provides a two-year warranty, as was the case for locally produced cars in the mid-nineties;
- provides a three-year warranty, as is currently the case for locally produced vehicles;
- provides a five-year warranty, as is the case for many imported cars;
- has a new product on the market for a limited amount of time of say, one to five years, with a warranty of at least the length of time of data observation.

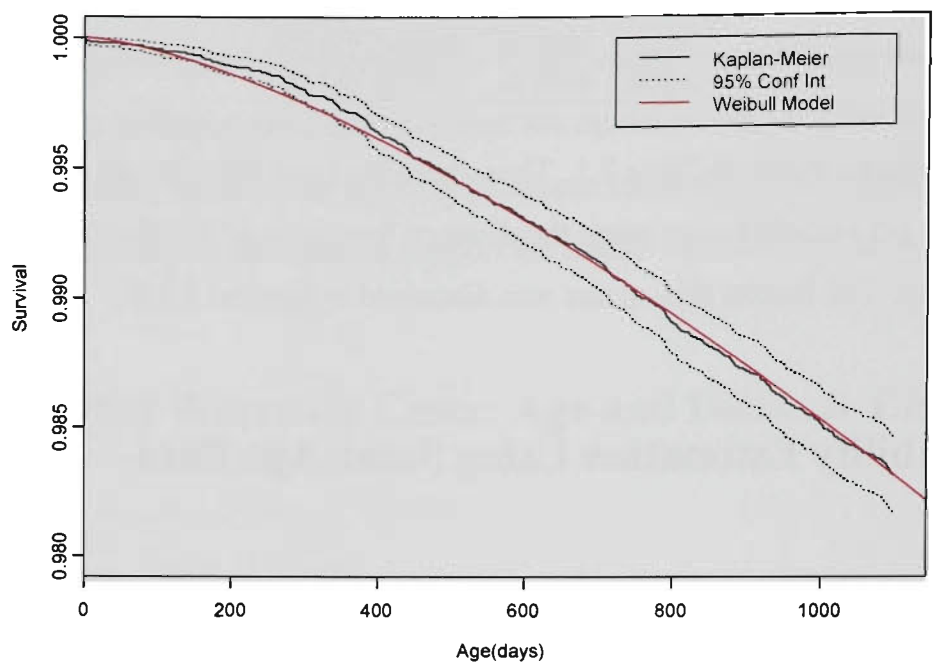


Fig. 7.3. *Kaplan-Meier and Weibull fits to the data.*

For the purpose of the simulations, the part "DISC,FT" was chosen as a typical part. The reliability of that part can be accurately modelled by the Weibull distribution, as shown in Figure 7.3, which shows both the Kaplan-Meier fit and the Weibull fit. The estimates of the Weibull parameters are $\widehat{\beta} = 1.462\,377$ and $\widehat{\eta} = 17\,788.75$. Although the simulations are set up with a specific part in mind, the results are equally valid for other parts, and for parts in other products, where the part is modelled by the Weibull distribution.

7.5.2 Method

Five simulations are conducted to generate failure times of a component for one year's production (30,138 vehicles), using the Weibull distribution parameters stated above. The simulations generate failure times of components that are all one, two, three, four or five years old. From the simulated data, the parameters of a Weibull model are estimated, as would be the case if they were unknown and data were collected over the stated periods.

In fitting a model to the simulated warranty data, the assumptions outlined in Section 5.2 still apply.

The warranty is taken to be three years. This limit applies to the vehicle, not the component. Thus, the warranty ends when a vehicle is three years old.

The simulation of failure times is conducted by the S-Plus function *fSmpPrms*, which is presented in Appendix E.3. The function generates a failure time for each component. If the failure time is less than the warranty period, a new failure time is generated and each vehicle's age is noted. In instances where the part fails again within the warranty period, a new failure time must again be generated until the vehicle is out of warranty. The product is observed for one to five years, in yearly increments, so the data will be censored at these times. From the simulated data, a database of failure and censored time is constructed, which is used to estimate the parameters of the Weibull model. This is done using the techniques described in the last two chapters.

Each simulation of warranty claims for a year's production is repeated 1,000 times for each observation period of one, two, three, four and five years. After each trial, the parameters of the Weibull distribution have been estimated. Thus, for each warranty period, 1,000 estimates of the Weibull parameters have been generated. The mean of each of the Weibull parameters, together with their variances and 95% confidence intervals, made up of the 2.5% to 97.5% quantiles, are calculated in the script *SimulDisc.ssc*, which can be found on the accompanying compact disc. The script calls the function *fSmpPrms,* and produces the relevant graphs.

7.5.3 Results

The means, variances and 95% confidence intervals of each parameter are shown in Table 7.3. The results are also presented in the boxplots of Figures 7.4 and 7.5. Clearly, the estimates from the one-year data are inadequate for both parameters. By two years, reasonable point estimates are obtained for both parameters. The variances are also quite reasonable. Further accuracy in parameter estimates is achieved, as would be expected, as the products' ages increase to five years.

The improvement in the parameter estimates can be clearly seen in Figures 7.6 and 7.7, which show the 3D density plots of the parameters over the observation times. The improvement from the two-year-old data over the one-year-old data is quite visible for both parameters. Both the point estimate and its variance show a marked improvement

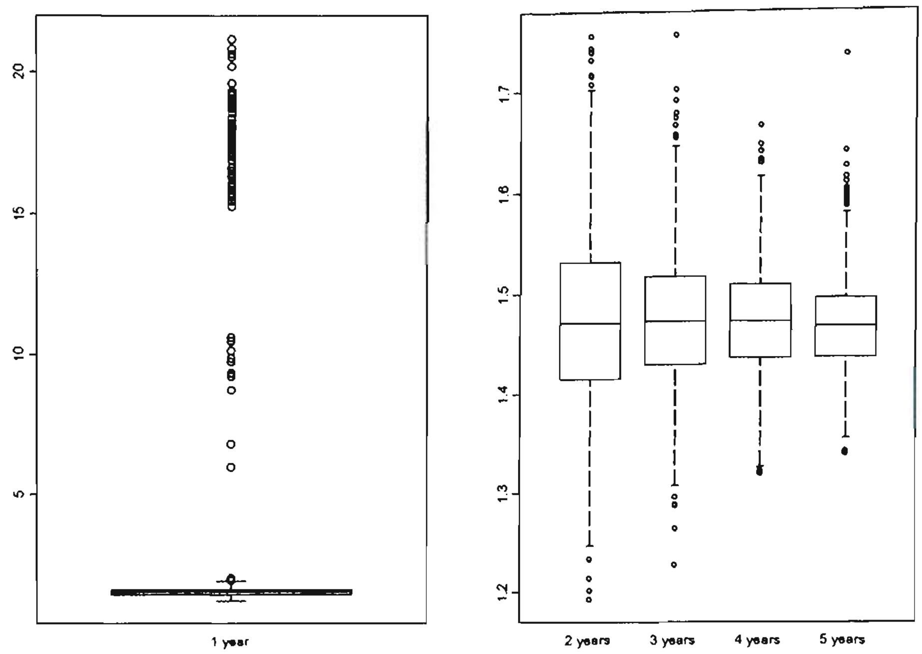


Fig. 7.4. $\widehat{\beta}$ *(shape parameter) boxplots.*

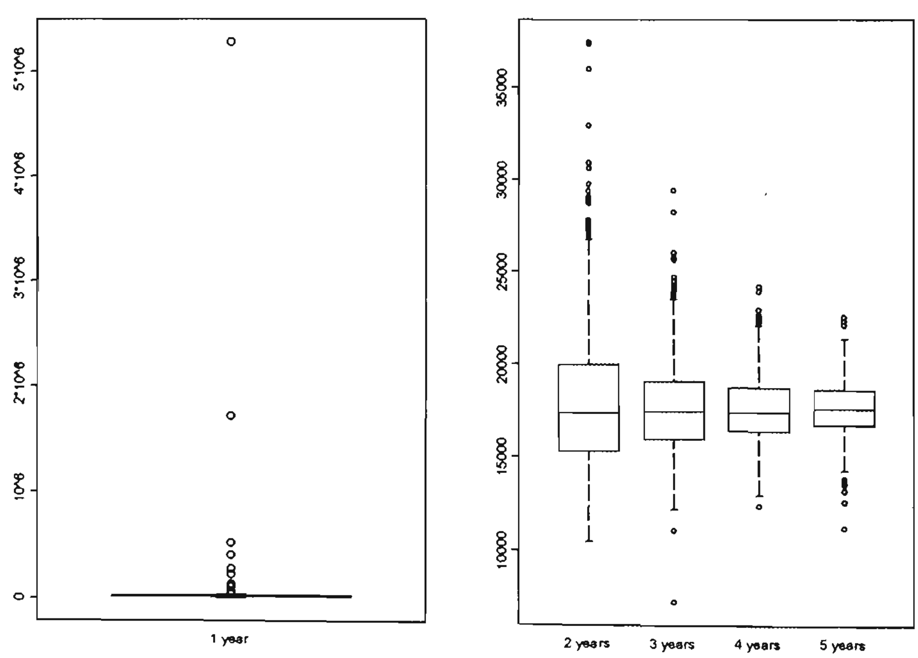


Fig. 7.5. $\widehat{\eta}$ *(scale parameter) boxplots.*

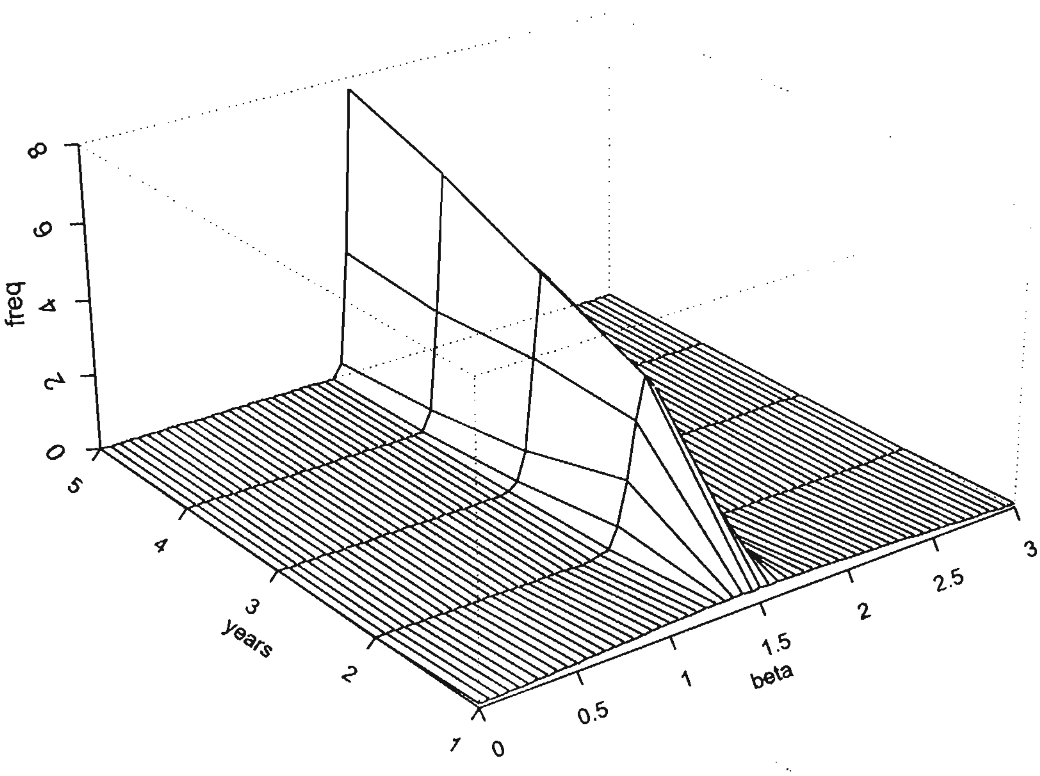


Fig. 7.6. $\widehat{\beta}$-time density plot.

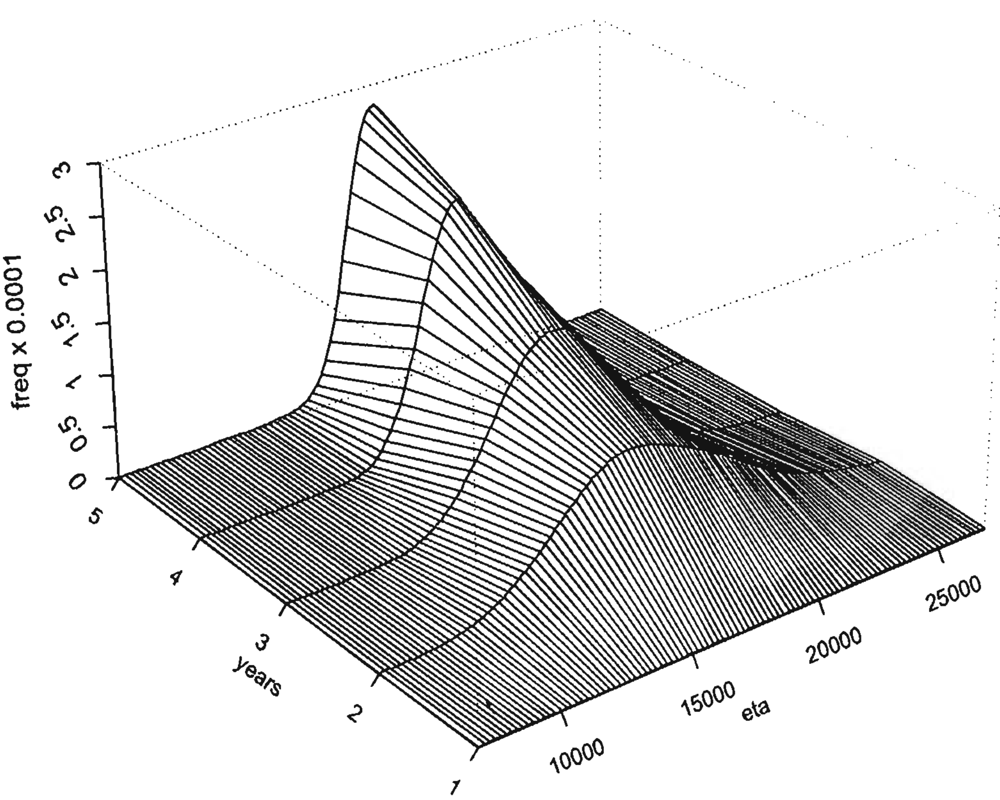


Fig. 7.7. $\widehat{\eta}$-time density plot.

Table 7.3. *Weibull parameter estimates.*

(a) β estimates.

Period	$\widehat{\beta}$	$\text{sd}(\widehat{\beta})$	95% CI for $\widehat{\beta}$
1 year	2.687 58	4.206 62	[1.235 61, 18.058 5]
2 years	1.474 12	0.088 314 7	[1.308 86, 1.662 96]
3 years	1.474 69	0.067 020 7	[1.349 65 1.616 30]
4 years	1.472 76	0.053 355 7	[1.370 73, 1.578 273]
5 years	1.467 57	0.046 569 9	[1.383 20, 1.562 23]

(b) η estimates.

Period	$\widehat{\eta}$	$\text{sd}(\widehat{\eta})$	95% CI for $\widehat{\eta}$
1 year	27 042.0	176 539	[3 349.347, 39 912.59]
2 years	17 913.2	3 679.70	[12 349.6, 26 397.3]
3 years	17 642.5	2 374.76	[13 531.2, 22 835.1]
4 years	17 628.3	1 681.56	[14 615.5, 21 422.5]
5 years	17 749.2	1 351.12	[15 345.5, 20 607.0]

The QQ-plots are used to determine whether the parameter estimates are normally distributed. These are shown in Figure 7.8. It can be seen that neither $\widehat{\beta}$ nor $\widehat{\eta}$ display normal behaviour for the one-year data. By two years, the data is approaching normality, confirming that we can obtain reasonable point and interval estimates for the parameters. The three, four and five year data displays normal behaviour. The QQ plots for five years are very similar to the four-year plots, and have been omitted for convenience.

7.5.4 Interpretation of Results

Weibull parameter estimates obtained from warranty data collected on one-year-old products are not very reliable. This is true of both the point and interval estimates of both Weibull parameters. A vast improvement in both point and interval estimates of the parameters is achieved if the product is two years old. Further accuracy is achieved with tighter confidence intervals by allowing the product to age further, but reasonable estimates can be obtained after just two years. These conclusions apply to products with a three-year warranty.

The above conclusions apply to the fitting of a Weibull model under the assumption that the Weibull model adequately represents the reliability of the component in the first place. Although this appears to be the case for the stated component, as is evident in Figure 7.3, the collection of further reliability data on that component may have revealed a variation from the model in the longer term. Therefore, conclusions beyond three years

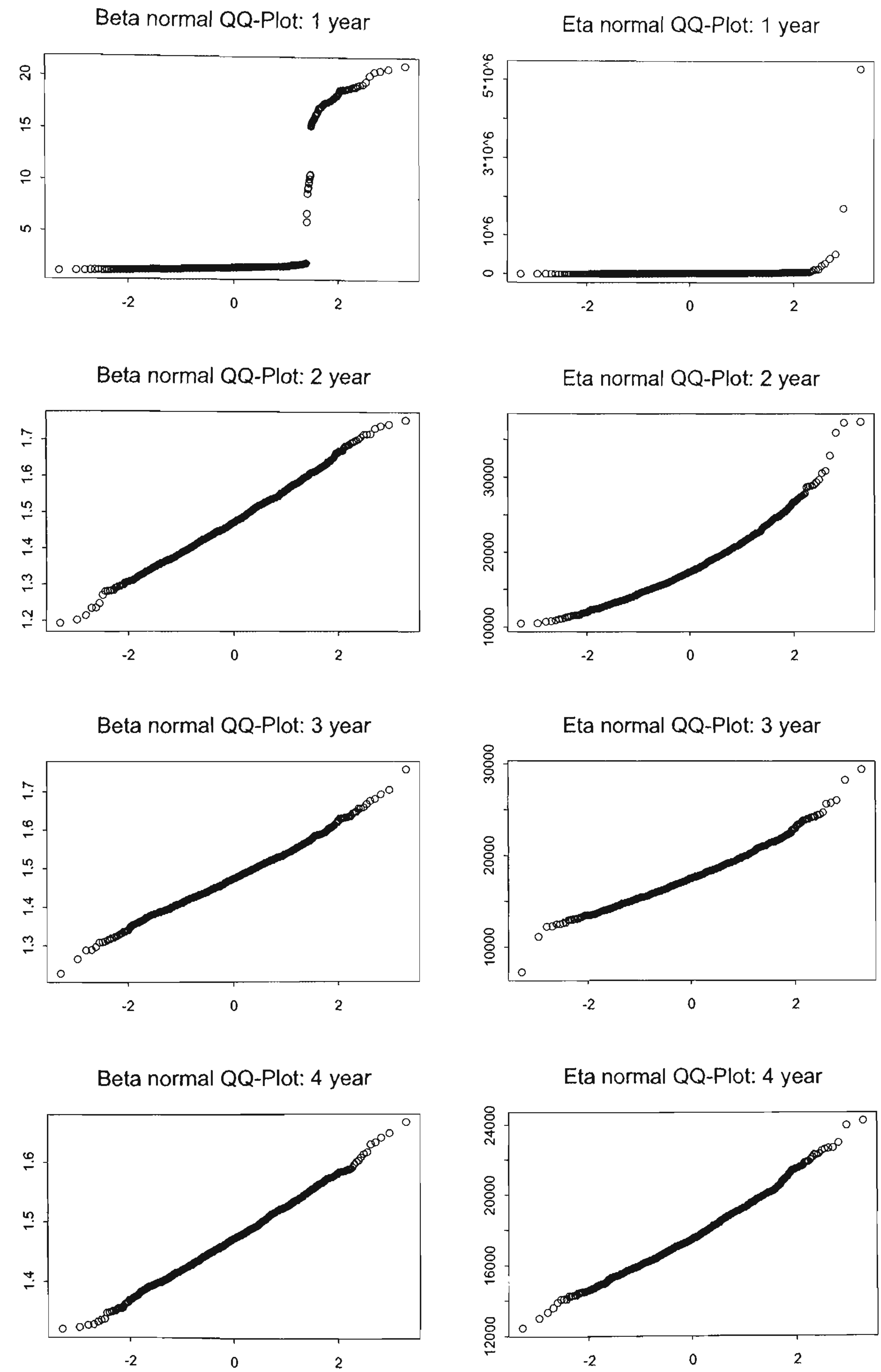


Fig. 7.8. $\widehat{\beta}$ *and* $\widehat{\eta}$ *QQ plots, showing sample quantiles (vertical axis) against normal quantiles (horizontal axis).*

for the stated component may not be valid, although they certainly apply if the Weibull model is a good representation of the reliability of the component.

Since most production occurs continuously over time rather than in yearly batches, it is informative to analyse warranty data continuously, rather than wait until all units have reached the required age. This is the subject of the simulations in the next section. However, the simulations of this section, although based on one-year data, may be applied to production that occurs in batches, such as between machinery adjustments or maintenance, or for runs with a constant component specification or supplier, where the time-frame would not usually be of one year's duration.

7.6 Reliability Estimation Using Varying-Age Data

7.6.1 Purpose and Background

The purpose of the simulations in this section is to test if, and when, an accurate estimate of the Weibull parameters can be made in a continuous-production environment when data is collected continuously over an observation period of two to five years.

In a continuous production environment, it is desirable to analyse warranty data before all units reach the end of their warranty period. In so doing, items will have varying ages at a particular point in time. So for example, after two years of production, the age of vehicles will range from zero to approximately two years. Simulations similar to the one of this section and Section 7.7 have been reported by Summit, Cerone and Diamond (2002b).

7.6.2 Method

The Weibull parameters used to simulate data in this section are the same as those used in the preceding section. Using these parameters, failure times are again simulated, as in the previous section. The difference in this simulation is that the vehicles will not all have the same age, as continuous production is being simulated. Thus, a production date, sale date and a component failure time are simulated. Data is generated for observation periods of two to five years in one year increments.

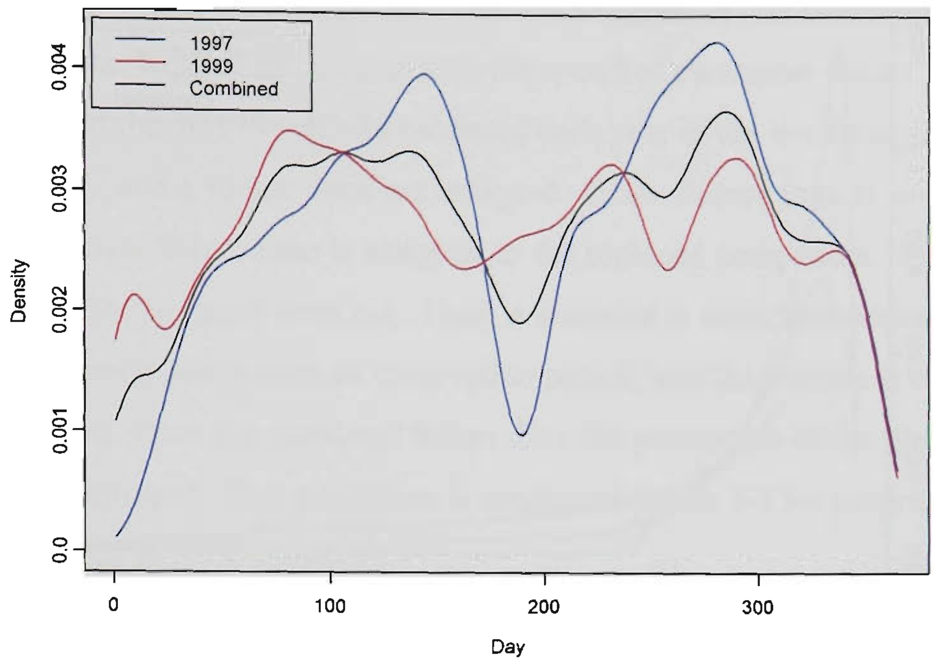


Fig. 7.9. *Daily production.*

In fitting a model to the simulated warranty data, the assumptions outlined in Section 5.2 have again been made. Yearly production has been set at 30,138 vehicles. The warranty period is again fixed at three years.

The simulations have again been repeated 1,000 times so that inferences on the distribution of estimated parameters can be made.

Simulating Production

The production figures of two one-year periods are used to determine a distribution of production dates throughout the year. Figure 7.9 shows the production density for the 1997 data used in the preceding chapter, together with 1999's production density, and the combined density. By using the combined density, any extremes in production that may be caused by occurrences such as machinery breakdowns, strikes, or the filling of large orders, have been smoothed out. However, seasonal influences affecting production, such as holiday breaks, have been retained, since they would be similar for the two production years. Sampling from this combined density is conducted to simulate a year's production. Production dates for subsequent years have then been generated by adding multiples of 365 to another set of simulated production days.

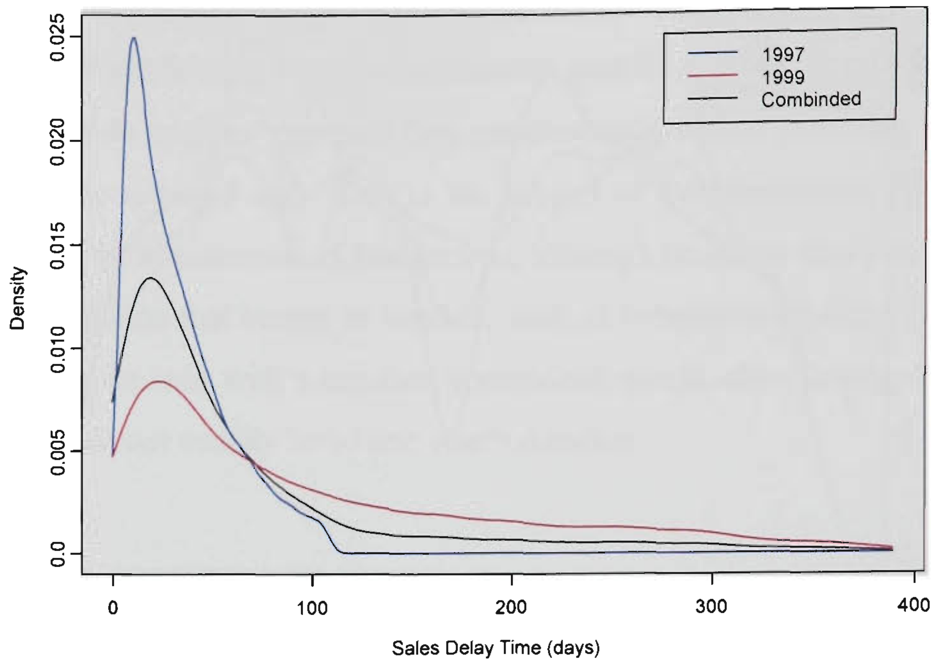


Fig. 7.10. *Sales delay time densities.*

The S-Plus function, *fProdDns* obtains the combined production density, and randomly generates production dates for any one-year period from the combined density. The script for the function is included in Appendix E.4 Although production varies according to the day of the week, (for example, lower production on weekends,) the function does not distinguish which day of the week a vehicle is produced, since that amount of detail is not required.

Simulating Sales

Products are not usually sold immediately after they are produced. There is a delay between production and sale. We shall refer to this time as the sale-delay time. Figure 7.10 shows the densities of sale-delay for the 1997 data from the preceding chapter, for the 1999 sale-delay times, and for the combined data. In generating these densities, errors and outliers have firstly been removed, as discussed previously in Subsection 4.4.2. It can be seen that the delay-time densities for the two years are very similar.

The S-Plus function *fDlayDns* generates sale-delay times from this density. The script for the function can be found in Appendix E.5.

Simulating Failure Times

As in the previous section, failure times are assigned in this simulation by random sampling from the Weibull distribution with the specified parameter values.

For each of the 30,138 vehicles produced each year in the simulation, a production day, a sale-delay and a failure time are assigned. If the failure time is within the warranty period, a new failure time is assigned to the replaced component. This process is continued until the warranty runs out. Thus, a situation is simulated where a product is manufactured continuously over an observation period, and failure data is collected over a warranty period. From the simulated failure data, the parameters of the Weibull reliability model are estimated. This procedure is conducted by the S-Plus program *fSmpCont*, which is presented in Appendix E.6.

The simulations are repeated 1,000 times for each observation period of two, three, four and five years. After each trial, the parameters of the Weibull distribution are estimated. The script *SimulCont.ssc*, which can be found on the accompanying compact disc, calculates the mean, standard deviations and a 95% confidence intervals (2.5 to 97.5 percentile range) of the Weibull parameters. It also produces boxplots, density and QQ plots. The script is also used to call the function *fSmpCont* and store the results of the simulations.

7.6.3 Results

The means, standard deviations and 95% confidence intervals, of both Weibull parameters, are shown in Table 7.4. It can be seen from Table 7.4(a) that the β estimates are biased downwards from the true value of 1.462 377. Similarly, Table 7.4(b) shows that the η estimates are biased upwards from their true value of 17,788.75. The combined effect of the two biases is to decrease the number of initial failures and increase the number of later failures. This can be clearly seen in Figure 7.11, which shows the Weibull probability density functions obtained by using the parameter estimates for 2 to 5 years from Table 7.4.

Table 7.4. *Weibull parameter estimates with continuous production.*

(a) β estimates.

Period	$\widehat{\beta}$	sd($\widehat{\beta}$)	95% CI for $\widehat{\beta}$
2 years	0.879 805	0.0415940	[0.799 686, 0.963 523]
3 years	0.894 480	0.025 492 8	[0.844 143, 0.943 817]
4 years	1.093 72	0.027 276 2	[1.039 13, 1.148 74]
5 years	1.190 90	0.025 273 3	[1.141 33, 1.238 74]

(b) η estimates.

Period	$\widehat{\eta}$	sd($\widehat{\eta}$)	95% CI for $\widehat{\eta}$
2 years	506 946	163 697	[265 508, 901 586]
3 years	320 725	54 758.1	[233 538, 452 948]
4 years	82 281.3	9 178.06	[65 931.2, 102 365]
5 years	50 284.6	4 238.61	[426 82.0 59 397.5]

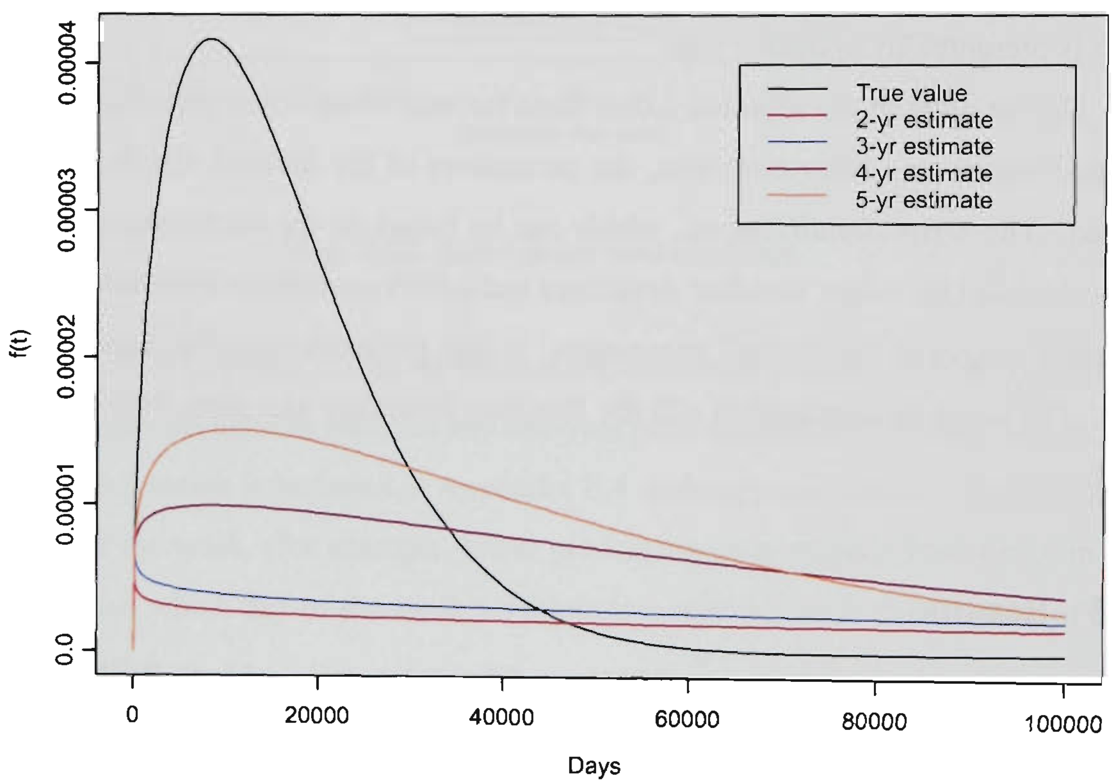


Fig. 7.11. *Probability density function for the estimated β and η values.*

Because the simulation produces vehicles continuously, there is a larger number of young vehicles than would be the case if the sample were not continuously added to. Thus, a large sample of early-censored data has been generated. If these early-censored items were observed for a longer time period, almost all would be failure-free data. Their existence in the survival database inflates the number at risk at an early age. The result is a reduction of the probability of early failure, and an increase in the probability at later ages. Thus, the failure distribution is biased downwards with a resulting bias in the parameter

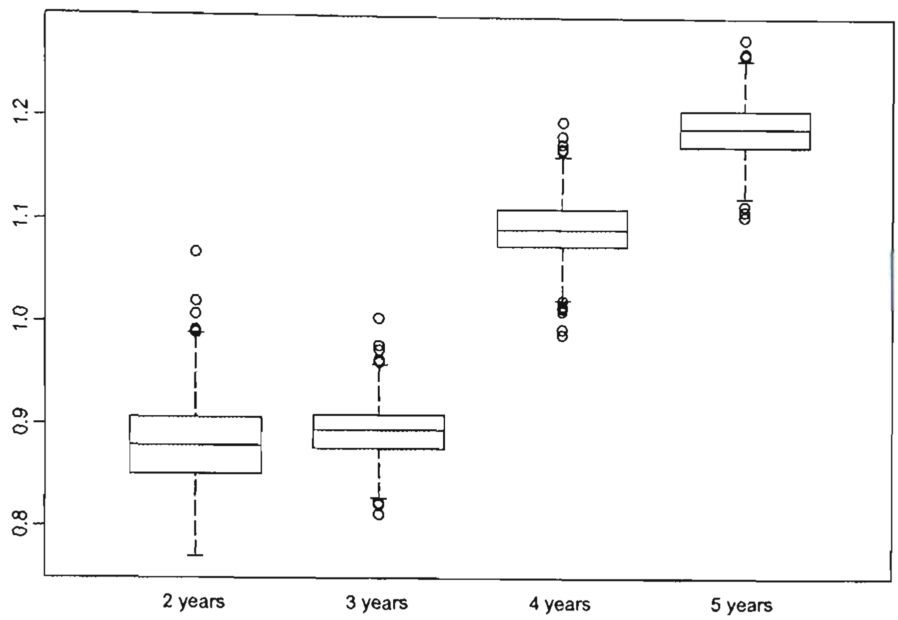


Fig. 7.12. $\widehat{\beta}$ *(shape parameter) boxplots using continuous data.*

estimates that reflects this underestimation. The simulation has been repeated using the exponential distribution instead of the Weibull, and the underestimation of the failure rate parameter has been evident there as well.

The boxplots for the same data are shown in Figures 7.12 and 7.13. The figures clearly show that there is quite a bias in the estimates obtained from the two- and three-year data, with a smaller, but still existent, bias by five years. There is a jump in the values of the parameter estimates at four years because by this time, most of the vehicles from the first year of production have reached the end of their warranty period, this being three years. Further improvement can be seen in the five-year data, because more vehicles have reached the end of the warranty period, and have been censored at that time. As more and more vehicles reach the end of the three-year warranty period, the influence of the bias in the earlier data decreases. However, the bias still exists at four years and beyond because production is continuous, and recently produced vehicles are still being censored at a young age.

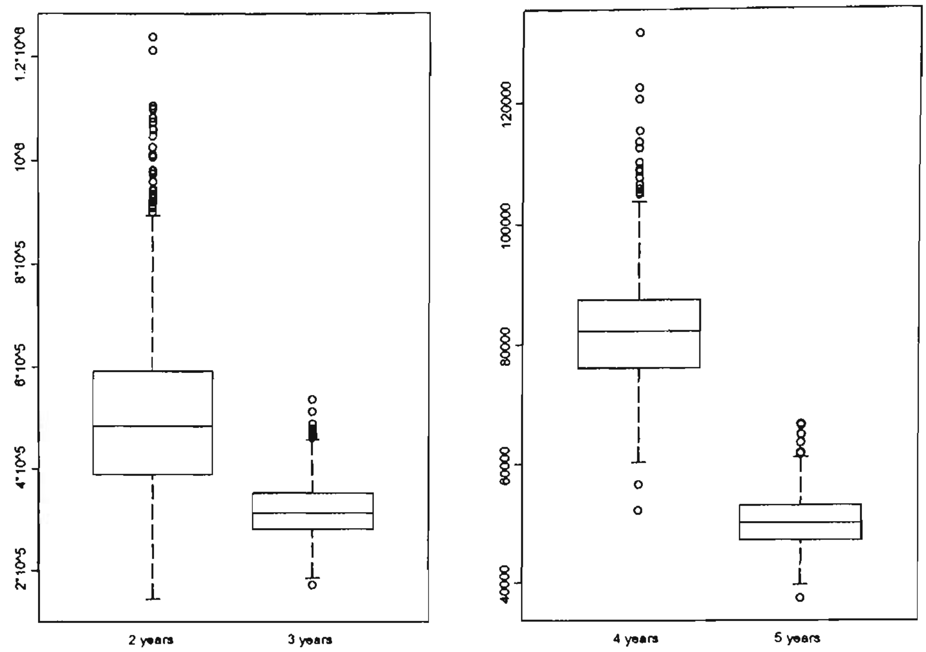


Fig. 7.13. $\widehat{\eta}$ *(scale parameter) boxplots using continuous data.*

7.7 Reliability Estimation Using Varying-Age Data From 1 Year's Production

7.7.1 Purpose

The purpose of the simulations in this section is to consider whether or not the biases that exist in the parameter estimates obtained from the simulations of the previous section can be eliminated, or reduced, when the data is restricted to just one year's production. Thus, newer production, which results in more early censored items, is deliberately excluded.

7.7.2 Method

The method used for the simulations in this section is the same as the one used in the previous section, except that failure data for only one year's production is generated. This single year's production is followed for observation periods of two, three, four and five years, but new production data is not added to the database.

As in the previous two sections, data from a Weibull distribution with $\beta = 1.462\,377$ and $\eta = 17\,788.75$, representing failure data for the part "DISC,FT", are simulated. Data

Table 7.5. *Beta estimates for one year's production.*

(a) β estimates.

Observation Time	$\hat{\beta}$	sd($\hat{\beta}$)	95% CI for $\hat{\beta}$
2 years	1.007 86	0.052 793 7	[0.909 816, 1.109 00]
3 years	1.127 57	0.045 198 3	[1.045 24, 1.222 30]
4 years	1.467 98	0.065 222 5	[1.338 40, 1.601 21]
5 years	1.471 42	0.065 353 9	[1.343 94, 1.603 69]

(b) η estimates.

Observation Time	$\hat{\eta}$	sd($\hat{\eta}$)	95% CI for $\hat{\eta}$
2 years	131 261	39 382.6	[76 709.3, 228 197]
3 years	55 309.8	9 153.60	[40 124.5, 75 750.9]
4 years	17 858.6	2 304.15	[13 931.0 22 838.6]
5 years	17 715.9	2 276.49	[13 946.7, 22 688.6]

representing one year's production of 30,138 are generated, from which the parameters of the Weibull model are estimated. This is done for each observation period of two to five years. The simulations are then repeated 1,000 times, so that the distributions of the parameters approach the normal distribution. This procedure is conducted in the S-Plus script *SimulCont1yr.ssc*, which is presented in Appendix E.7. The function that generates the sample, *fSmpCont1yr*, is included on the accompanying compact disc.

7.7.3 Results

The mean, variance and 95% confidence interval (2.5% to 97.5% quantile range) for the thousand simulations for each observation period are shown in Table 7.5.

The data are also displayed in the boxplots of Figures 7.14 and 7.15. The bias that can be seen in the two and three-year plots is not evident in the four and five-year plots. By four years, most of the vehicles would have reached the end of their warranty periods. Little is gained by continuing the observation to five years, as the data is censored at three years, that being the end of the warranty period. The slight gain is due to the fact that there are a few later selling vehicles, which would have reached the end of their warranty by five years.

Comparing Tables 7.4(a) and (b) with Tables 7.5(a) and (b), or alternatively Figures 7.12 and 7.13 with Figures 7.14 and 7.15, it can be seen that the bias in each observation period has been reduced by limiting the simulation to the observation of just one year's production. It can also be seen that by the fourth year of observation, the bias in the parameter estimates is no longer evident. This is a result of the maturation of the data to

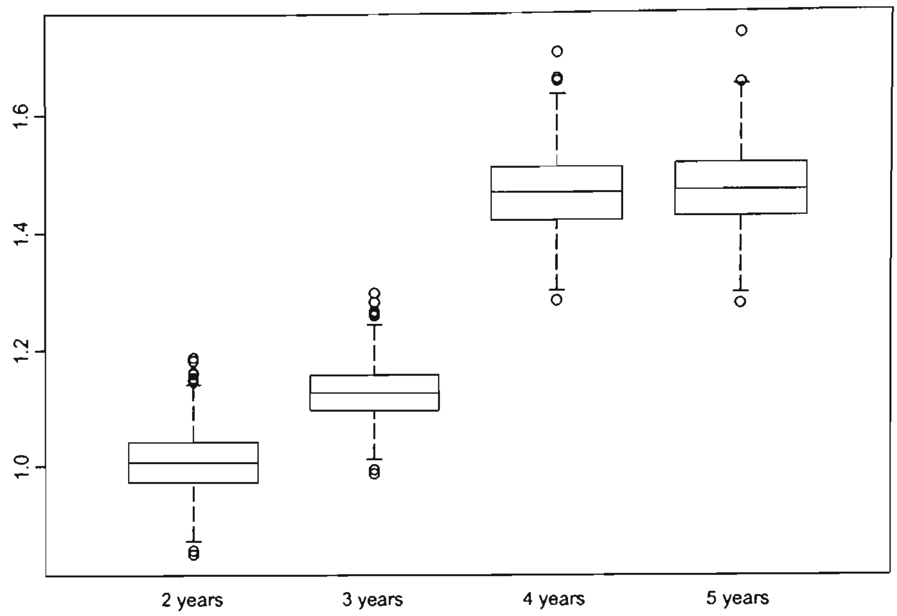


Fig. 7.14. $\widehat{\beta}$ (shape parameter) boxplots using 1 year production data.

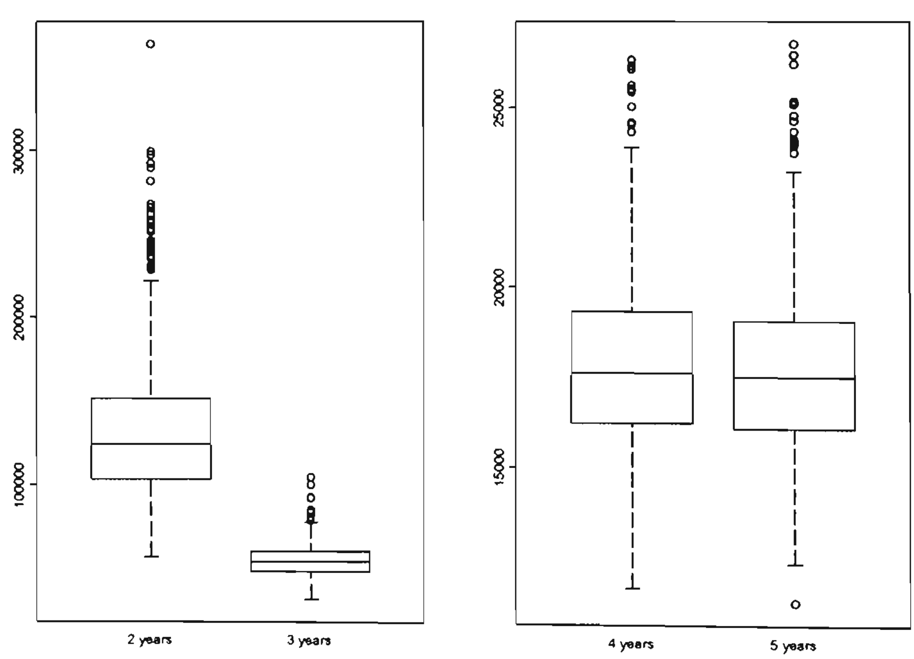


Fig. 7.15. $\widehat{\eta}$ (scale parameter) boxplots using 1 year production data.

its warranty limit of three years. As explained above, by limiting the number of early-censored data, the sample size is not allowed to become artificially inflated in the earlier periods.

7.7.4 Interpretation of Results

The results of the simulations in this and the previous sections show that a better estimate of reliability of a product can be obtained by limiting the amount of data to be used to estimate the parameters of the lifetime distribution to a set amount, say one year's production, and then by allowing that data to mature.

7.8 Conclusion

The results from the bootstrap (non-parametric) and parametric simulations covering all component failures support the results obtained by the modelling in the last chapter. Thus, it can be concluded that if components are modelled in the way they have been in the last chapter, that we could expect results for the warranty cost that are consistent with the observed results.

The last three simulations of this chapter have shown that reasonable Weibull-parameter estimates can be obtained from warranty data of products that are all two years old or more. If data is continuously being added to a database, so that the age of the items varies from zero up to the end of the observation period, then a bias is introduced into the estimates. Better results can be obtained by observing just one year's production for a given length of time, rather than by using additional new data. When continuous-production data is used, good parameter estimates can be obtained using an observation period of four years. By this time, all of the year's production have reached the end of their three-year warranty. Thus, the simulations validate the technique used in the previous chapter of using one year's production data, and observing that data to the end of its three-year warranty, to estimate the parameters of the Weibull models of component reliability.

From the last three simulations, some issues worthy of further investigation arise that go beyond the scope and focus of this study. One would expect a larger sample size to produce better parameter estimates than a smaller one. However, it has been seen that adding young-censored data introduces a bias in the parameter estimates. The question arises as to where the balance is between sample size and maturity of the data, that produces the most accurate parameter estimates. Looking at the older data, one question that could be considered concerns the bias in parameter estimates and whether or not they could be reduced if all data were to be censored at a pre-determined age, thus providing a smaller range in product ages.

Chapter 8
Concluding Remarks and Future Directions

8.1 Introduction

In conclusion, the findings of this study are summarised, and the limitations of the study are restated. An indication of some extensions to this study are also provided.

This study has modelled the cost of an automobile warranty and has used the model to estimate the cost of an extended warranty. Having evaluated the parameters of the models based on the manufacturer's data, it is possible to estimate the cost of any length of warranty using these parameters. This has been demonstrated in Chapter 6 for the five-year warranty. The simulations developed in Chapter 7 provide a means of estimating the warranty cost when both the time and distance limitations of the warranty are altered. Specific parameter estimates have been obtained from the manufacturer's data, which could be used for reliability analysis of components. The techniques developed in this study could be applied to other automobile manufacturers' data. Indeed, it is possible to use the techniques to estimate the cost of a warranty of other manufactured goods. The methods used in this study would be useful to a manufacturer who wishes to estimate the warranty cost when the terms of the warranty are changed.

8.2 Motivation For This Study

It was pointed out in Chapter 1 that the cost of servicing a warranty is a significant portion of the cost of production. Thus, there is a need to obtain a reasonable estimate of a warranty cost, and to allow for this cost in a warranty reserve.

Many different warranty-cost models have been developed in the literature, covering various aspects that affect warranty-cost modelling, and these have been reviewed in Chapters 2 and 3. However, as discussed in Chapter 1, a number of authors have expressed the need to apply these models to real data. It has been difficult to find such studies because of the proprietary nature of warranty data. This study has made available an Australian car manufacturer's warranty database covering one whole year's production over its entire warranty period, and has provided an extensive analysis of that data.

Australian warranty data has not been found in the literature previously. Of the four car manufacturers in Australia, two showed no interest in sharing their warranty data, one showed interest in any findings, but was reluctant to supply any data, and only one was willing to make the data available. A similar situation existed with car importers, where none of the importers that were approached showed a willingness to co-operate in a study of their warranty data. Thus, this study has only been possible because of the willingness of this one manufacturer. The identity of the manufacturer has not been divulged in order to protect the manufacturer from any misuse of statistics derived from the data.

8.3 Findings of This Study

It was observed from inspection and analysis that the data needed to be cleaned before it could be used to make inferences. Consequently, extensive techniques for checking and cleaning the data have been developed in this study, as discussed in Chapter 4. A number of custom-written S-Plus functions have been created in this study to conduct the necessary data cleaning. In order to conduct survival analysis with any accuracy, it has been important to obtain a true measure of a component's age. This has primarily been done from the recorded sale and repair dates, but has then been verified against the age and distance fields recorded in the database. The result of checking and cleaning the manufacturer's data has been the creation of a data set that could be confidently analysed. Although the data checking and cleaning in this study has been extensive, it should be pointed out that less than two percent of the claims data were in error.

After the data were cleaned, they were used both to calculate components' ages and to create a survival database of component failure and censored times during the warranty period. The details of this process have been discussed in Chapter 4.

Warranty-cost models that can be applied to survival data have been discussed in Chapter 5. Modelling the warranty process as a renewal process, the expected number of replacements during the warranty period has been estimated and used to obtain an estimate of the warranty cost. Numerical methods have been employed to determine the expected number of renewals for the Weibull distribution. Warranty-cost models based on the expected number of renewals have been developed to obtain point estimates and their variances. Various parametric models have been considered, with the log likelihood values of all models being fairly close to one another, as reported in Chapter 6.

The total warranty cost per vehicle has then been modelled based on the sum of the individual components' warranty costs. This, of course, assumes independence amongst component failures, a necessary assumption to be able to analyse the large number of components included in the manufacturer's database. Other assumptions have been discussed in Section 5.2; the models developed in this study are subject to those assumptions being true.

In addition to obtaining point estimates for the warranty cost per vehicle, techniques for obtaining interval estimates have been discussed in Chapter 5. These have been based on the variance of the number of renewals and the variance of the cost of repair. The variance for the exponential model was readily evaluated, but that of the Weibull model relied on numerical methods for its calculation.

8.3.1 Comparison of Models

The exponential model provided the same point estimate of the warranty cost per vehicle for the current warranty of three years as did the Weibull model, as discussed in Chapter 6. In fact, the bootstrap model also obtained the same results, as reported in Chapter 7. Thus the additional processing required for the Weibull model is not deemed to be justified when estimating the cost of the current model.

For the extended warranty of five years, the exponential and Weibull models gave quite different results. It was reported in Chapter 6 that for some components, the parameter estimates of the Weibull model proved to be unrealistic. However, by removing one or two of the oldest failure times, a much better fit was able to be obtained. This "adjusted" Weibull model predicted much higher warranty costs for the extended warranty as compared with the exponential model. Extrapolation using the exponential model appeared to be somewhat dangerous, based on these results. Caution needs to be taken even when extrapolating with the Weibull model. Attempting to extrapolate beyond anything but a short distance from the observed points could lead to inaccurate predictions. This was evident for a number of components, where the extended-warranty costs based on the estimated Weibull models did not result in realistic costs predictions. Thus, the Weibull model was rejected for those components. It becomes increasingly clear that care needs to be taken when extrapolating beyond the warranty period when one considers that only a very small portion of a component's life is observed during the warranty period.

The exponential model provided a reasonable confidence interval on the warranty cost. However, poor, unusable confidence intervals were obtained using the Weibull model. This occurred because of the large variances in the expected number of renewals for most components. This in turn, was due to the fact that the interval estimates were obtained from a confidence region based on the confidence intervals of the two Weibull parameters. However, a workable confidence interval was able to be obtained using the variance of the exponential model.

Existing knowledge of the theoretical warranty cost models has been extended through the development of techniques that enable these models to be used with real data. Extensive computer coding, in the form of numerous S-Plus functions, has been developed in this study. This has been necessary as the existing S-Plus functions need to be supplemented to perform the modelling required in this study. The new functions are designed to transform the manufacturer's raw data into an estimate of the warranty cost.

8.3.2 Findings of the Simulations

In Chapter 7, simulations have been used to compare the expected warranty cost based on the estimated parameter values of the models, with the observed warranty cost. The simulation of warranty cost for a three-year warranty provided similar results to that of the observed warranty cost, thus supporting the parametric models developed in Chapter 6.

A second simulation used the parameters that were determined in Chapter 6, and imposed a distance limit on the warranty. The warranty costs for both the three-year and five-year warranties are not significantly different to the results obtained from the simulation that did not have the distance limitation. Thus, it can be seen that using the one-dimensional approach to modelling the warranty cost does not greatly affect the results.

Other simulations in Chapter 7 have shown that by collecting one year's warranty data until the end of the warranty period, an accurate model of a component's reliability can be obtained. The simulations in Chapter 7 have also shown that parameter estimates can be biased if data is not sufficiently mature, leading to erroneous warranty cost estimates.

8.4 Extensions to This Research

It may be informative to model the early failure data separately from the later data. Similarly, modelling the post-warranty claims as a separate group could also provide further information. By so doing, the characteristics of these stratified groups could possibly be better catered to.

Conversely, one may ask if the modelling of components separately, as has been done in this study, is worth the greater expenditure in analysis and computation time. The techniques used in this study have been expensive in terms of both the time taken to set up the analysis and the processing time required to do the analysis. Linear extrapolation of monthly warranty claims figures is the most common approach used by manufacturers. It would be informative to compare the results of this approach with those obtained by the techniques outlined in this study. It would also be worthwhile to compare the warranty cost obtained from modelling the reliability of subsystems, with that obtained by modelling the reliability of components separately.

The reliability of parts obtained from different suppliers can be monitored through the analysis of warranty data. It may be worthwhile to analyse warranty data of components from different suppliers in order to compare their reliability. Warranty data analysis could be used as part of the quality control process. It would be informative to conduct a study that estimated the expected warranty savings of such analysis.

An area that requires further work is that of the interval estimation of the warranty cost per vehicle, obtained by using the Weibull model, as developed in Chapter 5. As has been stated in Chapter 6, the results obtained from this model are not usable. This is as a result of the large variances obtained for the expected number of renewals of most parts. The modifications that can be made to improve the interval estimates need further investigation.

In the simulations used in this study, data has been generated from a set of fixed parameters. The values of the parameters have been estimated from the warranty data. However, these values are not known exactly, and this fact could be incorporated into the simulation. This would be achieved by firstly sampling the parameter values, and then generating simulated data using these sampled values. By so doing, the uncertainty in the parameters would be reflected in the simulated data. This would produce a greater variation in the warranty costs, as seen in the real data.

It is intended that future studies by the author address some of these areas of further investigation.

8.5 Conclusion

In conclusion, this thesis has used the exponential and Weibull distributions to model the reliability of components of a particular model of car using the manufacturer's warranty data. This has been done after comparing these models with various other models, which all produced similar results. For each component, the expected number of renewals was estimated using the exponential and Weibull models, which, together with the estimated cost of repairing the component, was used to estimate the cost of the warranty. The two models were used to estimate the cost of an extended warranty, and comparisons were made between the models.

Simulations have been used to compare the predicted warranty cost from the models with the actual warranty cost. They were also used to validate some of the assumptions made, and the techniques used, in deriving the models for warranty cost.

The reliability models that have been established in this study are specific to a particular model from a specific manufacturer. The models would be very useful to this manufacturer in determining if variations from this reliability are evident in the manufacturing process when certain events occur, such as a change of supplier, batch lot, or a change in the manufacturing process. It is anticipated the findings of this study will be used in that fashion. Although the results are specific to this manufacturer, the techniques that have been developed are useful to other automobile manufacturers, and indeed to other complex products made up of a several components.

References

Amato, H.N. and Anderson, E.E. (1976), Determination of warranty reserves: an extension. *Management Science*, **22**, 1391-1394.

Ascher, H. (1992), Basic probabilistic and statistical concepts for maintenance of parts and systems. *IMA Journal of Mathematics Applied in Business and Industry*, **3**, 153-167.

Ascher, H. and Feingold, H. (1984), *Repairable Systems Reliability Modelling, Inference, Misconceptions and Their Causes*. Marcel Dekker, Inc., New York.

Baker, C.T.H. (1977), *The Numerical Treatment of Integral Equations*. Clarendon Press, Oxford.

Balachandran, K.R., Maschmeyer, R.A. and Livingstone, J.L. (1981), Product warranty period: a Markovian approach to estimation and analysis of repair and replacement costs. *The Accounting Review*, **56**, 115-124.

Baxter, L.A., Scheuer, E.M., Blischke, W.R and McConalogue, D.J. (1981), *Renewal Tables*. Graduate School of Business Administration and School of Business, University of Southern California.

Beerworth, E.E. (1991), Product liability in the 90s: The good news and the bad news for manufacturers. *Contemporary Issues in Product Liability Law,* E.E. Beerworth (Editor). The Federation Press, Sydney, 9-44.

Berke, T.M. and Zaino, N.A. (1991), Warranties: What are they? What do they really cost. *1991. Proceedings Annual Reliability and Maintainability Symposium*, 326-331.

Blischke, W.R. (1990), Mathematical models for analysis of warranty policies. *Mathematical and Computer Modelling*, **13**, 1-16.

Blischke, W.R. and Murthy, D.N.P. (1992), Product warranty management - I: A taxonomy for warranty policies. *European Journal of Operational Research*, **62**, 127-148.

Blischke, W.R. and Murthy, D.N.P. (1994), *Warranty Cost Analysis*. Marcel Dekker, Inc, New York.

Blischke, W.R. and Murthy, D.N.P. (2000), *Reliability: Modelling, Prediction, and Optimisation*. John Wiley and Sons, New York.

Blischke, W.R. and Scheuer, E.M. (1975), Calculation of the cost of warranty policies as a function of estimated life distributions. *Naval Research Logistics Quarterly*, **22**, 681-696.

Blischke, W.R. and Scheuer, E.M. (1981), Application of renewal theory in analysis of the free-replacement warranty. *Naval Research Logistics Quarterly*, **28**, 193-205.

Bohoris, G.A., and Yun, W.Y. (1995), Warranty costs for repairable products under hybrid warranty. *IMA Journal of Mathematics Applied in Business and Industry*, **6**, 13-24.

Bowerman, B.L. and O'Connell, R.T. (1990), *Linear Statistical Models: An Applied Approach* (2nd Edition). Duxbury, Pacific Grove, California.

Chen, T. and Popova, E. (2002), Maintenance policies with two-dimensional warranty. *Reliability Engineering and System Safety*, **77**, 61-69..

Chen, J., Yao, D.D. and Zheng, S. (1998), Quality Control for products supplied with warranty. *Operations Research*, **46**, 107-115.

Christer, A.H. (1981), The purposeful delay of maintenance work: economic and behavioural aspects, *Applied Systems and Cybernetics: Proceedings of the International Congress on Applied Systems Research and Cybernetics*, G.E. Lasker (Editor). Pergamon Press.

Christer, A.H. and Waller, W.M. (1984), Reducing production downtime using delay time analysis. *Journal of Operational Research Society*, **35**, 499-512.

Chun, Y.H. (1992), Optimal number of periodic preventive maintenance operations under warranty. *Reliability Engineering and System Safety*, **37**, 223-225.

Chun, H.C. and Tang, K. (1999), Cost analysis of two-attribute warranty policies based on the product usage rate. *IEEE Transactions on Engineering Management*, **46,** 201-209.

Cook, R.D. and Weisberg, S. (1994), *An Introduction to Regression Graphics*. John Wiley and Sons, New York.

Cox, D.R. (1982) *Renewal Theory*. Chapman and Hall, New York.

Cox, D.R. and Oakes, D. (1984), *Analysis of Survival Data*. Chapman and Hall, London.

Crow, L.H. (1990), Evaluating the reliability of repairable systems. *1990 Proceedings Annual Reliability and Maintainability Symposium*, IEEE, 275-278.

Crow, L.H. (1993), Confidence intervals on the reliability of repariable systems. *1990 Proceedings Annual Reliability and Maintainability Symposium*, IEEE, 126-134.

Dagpunar, J.S. (1992), A comparison of adaptive and stationary repair-cost limit criteria for a producer's warranty. *IMA Journal of Mathematics Applied in Business and Industry,* **4**, 273-278.

Draper, N.R. and Smith, H. (1981), *Applied Regression Analysis*. John Wiley & Sons, New York.

Efron, B. and Tibshirani, R.J. (1993), *An Introduction to the Bootstrap*. Chapman and Hall, New York.

Eliashberg, J. Singpurwalla, N.D. and Wilson, S.P. (1997), Calculating the reserve for a time and usage indexed warranty. *Management Science*, **43**, 966-975.

Ellis, D.E. (1990), Extended service coverage cost predictions. *1990 Proceedings Annual Reliability and Maintainability Symposium*, IEEE, 233-236.

Elsayed, E.A. (1996), *Reliability Engineering*. Addison Wesley Longman, Massachusetts.

Faddy, M.J. (1993), Compartmental models of times to failure, *Proceedings of the Australia-Japan Workshop on Stochastic Models in Engineering, Technology and Management*, S. Osaki and D.N.P Murthy (Editors). World Scientific, Singapore, 134-142.

Farewell, V. T. and Cox, D. R. (1979), A note on multiple time scales in life testing. *Applied Statistics*, **28**, 73-75.

Fleming, T.R. and Harrington, D.P. (1984), Nonparametric estimation of the survival distribution in censored data. *Communications in Statistics*, **13**, 2469-2486.

Frees, E.W. (1986), Warranty analysis and renewal function estimation. *Naval Research Logistics Quarterly*, **33**, 361-372.

Frees, E.W. (1995), Semiparametric Estimation of Warranty Costs. *Nonparametric Statistics*, **5**, 103-121.

Frees, E.W. and Nam, S.H. (1988), Approximating Expected Warranty Costs. *Management Science*, **34**, 1441-1449.

Gertsbakh, I.B. and Kordonsky, K.B. (1998), Parallel time scales and two-dimensional manufacturer and individual customer warranties. *IIE Transactions*, **30**, 1181-1189.

Gilbertson, R. J. (1989), Risk reduction through application of essential performance requirements warranties. *Proceedings fo the IEEE 1989 National Aerospace and Electronics Conference*, IEEE. New York, 2010-2014.

Giuntini, R.E. and Giuntini, M.E., (1993), Simulating a Weibull posterior using Bayes inference, *1993 Proceedings Annual Reliability and Maintainability Symposium*, 48-55.

Hadel, J.J. and Lakey, M.J. (1993), A structured approach to warranty preparation and risk assessment. *1993 Proceedings Annual Reliability and Maintainability Symposium*, 298-304.

Hill, V.L., Beall, C.W. and Blischke, W.R. (1991), A simulation model for warranty analysis, *International Journal of Production Economics*, **22**, 131-140.

Hill, V.L. and Blischke, W.R., (1987), An assessment of alternative models in warranty analysis. *Journal of Information and Optimization Sciences*, **8**, 33-55.

Hu, X.J., Lawless, J.F. and Suzuki, K. (1998), Nonparametric estimation of a lifetime distribution when censoring times are missing. *Technometrics*, **40**, 273-285.

Hyundai (2004), Timeline [Internet.], Available at: http://www.hyundai.com.au/company_timeline.asp, accessed 22/6/2004.

Isaacson, D.N., Reid, S. and Brennan, J.R. (1991), Warranty cost-risk analysis. *1991 Proceedings Annual Reliability and Maintainability Symposium*, 332-339.

Iskandar, B.P. (1991), A two-dimensional renewal function solver. Department of Mechanical Engineering, Research Report No. 4/91, The University of Queensland, Australia.

Iskandar, B.P. and Murthy, D.N.P. (2003), Repair-replace strategies for two-dimensional warranty policies. *Mathematical and Computer Modelling*, **38**, 1233-1241.

Jablonowski, M. (2003), Working models: from theory to practice. *Risk Management Analysis*, **50**, December 2003, 48.

Jardine, A.K.S. (1984), Engineering maintenance management short course lecture notes. Department of Engineering Management, Royal Military College of Canada.

Jiang, R. and Murthy, D.N.P. (1995), Reliability modelling involving two Weibull distributions, *Reliability Engineering and System Safety*, **47**, 187-198.

Jiang, R. and Murthy, D.N.P. (1999), The exponentiated Weibull family: A graphical approach, *IEEE Transactons on Reliability*, **48**, 68-72.

Jun, B. and Pham, H. (2004), Discouned warranty cost of minimally repaired series systems. *IEEE Transactions on Reliability*, **53**, 37-42.

Kalbfleisch, J.D. and Lawless, J.F. (1988), Estimation of reliability in field-performance studies. *Technometrics*, **30**, 365-378.

Kalbfleisch, J.D., Lawless, J.F. and Robinson, J.A. (1991), Methods for the analysis and prediction of warranty claims. *Technometrics*, **33**, 273-285.

Kar, T.P. and Nachlas, J.A. (1997), Coordinated warranty and burn-in strategies. *IEEE Transactions on Reliability*, **46**, 512-518.

Kelley, C.A. (1996), Warranty Legislation. *Product Warranty Handbook*, W.R. Blischke and D.N.P. Murthy (Editors). Marcel Dekker, Inc, New York.

Kim, C.S., Djamaludin, I. and Murthy, D.N.P. (2001), Warranty cost analysis with heterogenous usage intensity. *International Transactions in Operational Research*, **8**, 337-347.

Lai, C.D., Xie, M. and Murthy, D.N.P. (1998), The beta integrated model. *Proceedings International Workshop on Reliability Modelling and Analysis–From Theory to Practice*, 69-104.

Lai, C.D., Xie, M. and Murthy, D.N.P. (2003), A modified Weibull distribution. *IEEE Transactions on Reliability*, **52**, 33-37.

Lakey, M.J. (1991), Statistical analysis of field data for aircraft warranties. *1991 Proceedings Annual Reliability and Maintainability Symposium*, 340-344.

Lawless, J.F. (1982), *Statistical Models and Methods for Lifetime Data*. John Wiley and Sons, New York.

Lawless, J.F. (1995), The analysis of recurrent events for multiple subjects. *Applied Statistics*, **44**, 487-498.

Lawless, J.F. (1998), Statistical analysis of product waranty data. *International Statistical Reviews*, **66**, 41-60.

Lawless, J.F., Hu, J. and Cao, J. (1995), Methods for the estimation of failure distributions and rates from automobile warranty data. *International Statistical Reviews*, **66**, 41-60.

Lawless, J.F. and Kalbfleisch, J.D. (1992), Some issues in the collection and analysis of field reliability data, *Survival Analysis: State of the Art*, J.P. Klein and P.K. Goel (Editors). Dordrecht: Kluer Academic Press, Boston, 141-152.

Lawless, J.F. and Nadeau, C, (1995), Some simple robust methods for the analysis of recurrent events. *Technometrics*, **37**, 158-168.

Lin, D., Zuo, M., Yam, R. and Meng, M. (2000), Optimal system design considering warranty, periodic preventive maintenance, and minimal repair. *Journal of the Operational Research Society,* **7**, 869-874.

Loomba, A.P.S. (1996), Historical perspective on warranty. In *Product Warranty Handbook*, W.R. Blischke and D.N.P. Murthy, (Editors). Marcel Dekker, Inc, New York.

Lu, M. (1998), Automotive reliability prediction based on early field failure warranty data. *Quality and Reliability Engineering International*, **14**, 103-108.

Lu, M., and Vance R.J. (1997), A distribution on multiple repairs per vehicle. *Quality Engineering*, **10**, 255-261.

Majeske, K.D. and Herrin, G.D., (1995), Assessing mixture-model goodness-of-fit with an application to automobile warranty data. *1995 Proceedings Annual Reliability and Maintainability Symposium*, 378-383.

Majeske, K.D. and Herrin, G.D., (1998), Determining the warranty benefits for automobile design changes. *1998 Proceedings Annual Reliability and Maintainability Symposium*, 94-99.

MathSoft (1999), *The S-Plus 2000 Guide to Statistics, Volume 2*. Data Analysis Products Division, MathSoft Inc, Seattle, Washington.

Martz, H.F. and Waller, R.A., (1991), *Bayesian Reliability Analysis*, Krieger Publishing Company, Malabar, Florida.

Menezes, M.A,J. and Quelch, J.A. (1990), Leverage your warranty program. *Sloan Management Review*, **31**, 69-80.

Menke, W.W. (1969), Determination of warranty reserves. *Management Science*, **15**, 542-549.

Mitra, A., and Patankar, J.G. (1993), An integrated multicriteria model for warranty cost estimation and production. *IEEE Transactions on Engineering Management*, **40**, 300-311.

Mood, A.M., Graybill, F.A. and Boes, D.C. (1974), *Introduction to the Theory of Statistics* (Third Edition). McGraw-Hill, New York.

Morgan, F.W. (1982), Marketing and product liability: A review and update. *Journal of Marketing*, **46**, (Summer, 1982) 69-78.

Moskowitz, H., and Chun, H.C. (1994), A Poisson regression model for two-attribute warranty policies. *Naval Research Logistics*, **41**, 355-376.

Moskowitz, H., and Chun, H.C. (1996), Two-dimensional free-replacement warranties. *Product Warranty Handbook, W.R.* Blischke and D.N.P. Murthy (Editors). Marcel Dekker, Inc, New York. 341-363.

Murthy, D.N.P. (1992), A usage dependent model for warranty costing. *European Journal of Operational Research*, **57**, 89-99.

Murthy, D.N.P. and Blischke, W.R. (1992a), Product warranty management – II: An integrated framework for study. *European Journal of Operational Research*, **62**, 261-281.

Murthy, D N.P. and Blischke, W.R. (1992b), Product warranty management – III: A review of mathematical models. *European Journal of Operational Research*, **63**, 1-34.

Murthy, D.N.P. and Blischke, W.R. (2000), Strategic warranty management: A life cycle approach. *IEEE Transactions on Engineering Management*, **47**, 40-54.

Murthy, D.N.P. and Djamaludin, I. (2002), New product warranty: A literature review. *International Journal of Production Economics*, **79**, 231-260.

Murthy, D.N.P. and Iskandar, B.P. (1992), One-dimensional renewal function solver, Report 1/92, Department of Mechanical Engineering, University of Queensland.

Murthy, D.N.P., Iskandar, B.P. and Wilson, R.J. (1995), Two-dimensional failure-free warranty policies: Two-dimensional point process models. *Operations Research*, **43**, 356-366.

Nelson, W. (1982), *Applied Life Data Analysis*. John Wiley and Sons, New York.

Nenoff, L. (1988), Field operational R & M warranty, *1988. Proceedings Annual Reliability and Maintainability Symposium*, 227-230.

Oh, Y.S. and Bai, D.S. (2001), Field data analyses with additional after-warranty failure data. *Reliability Engineering and System Safety*, **72**, 1-8.

Pham, H. and Wang, H. (1996), Imperfect maintenance. *European Journal of Operational Research*, **94**, 425-438.

Polatoglu, H. and Sahin, I. (1998), Probability distributions of cost, revenue and profit over a warranty cycle. *European Journal of Operational Research,* **108**, 170-183.

Quesenberry, C.P. and Kent, J., (1982), Selecting among probability distributions used in reliability. *Technometrics*, **24**, 59-65.

Rai, B. and Singh, N. (2003), Hazard rate estimation from incomplete and unclean warranty data, *Reliability Engineering and System Safety*, **81**, 79-92.

Rai, B. and Singh, N. (2004), Modelling and analysis of automobile warranty data in the presence of bias due to customer-rush near warranty expiration limit. *Reliability Engineering and System Safety*, **86**, 83-94.

Roberts, W.T. and Mann, L. (1993), Failure predictions in repairable multi-component systems. *International Journal of Production Economics*, **29**, 103-110.

Rooker, T. (1989), Warranty program engineering. *1989 Proceedings Annual Reliability and Maintainability Symposium*, 9-14.

Sarawgi, N. and Kurtz, S.K. (1995), A simple method for predicting the cumulative failures of consumer products during the warranty period. *1995 Proceedings Annual Reliability and Maintainability Symposium*, 384-390.

Singpurwalla, N.D. and Wilson, S. (1993), The Warranty Problem: Its Statistical and Game Theoretic Aspects. *SIAM Review*, **35**, 17-42.

Soland, R.M. (1968), Renewal functions for Gamma and Weibull distributions with increasing hazard rates, *Technical Paper RACTP-329*, Research Analysis Corporation, McLean, Virginia.

Stahl, M.J. and Grigsby, D.W. (1997), *Strategic Management: Total Quality and Global Competition*, Cambridge, M.A., Blackwell.

Stern, H. S. (1996), Neural networks in applied statistics. *Technometrics*, **38**, 205-214.

Summit, R. and Cerone, P. (2003), Costing a New Car Warranty, *Asia-Pacific Journal of Operational Research*, **20**, 89-101.

Summit, R.A. Cerone, P. and Diamond, N.T. (2002a), Simulating reliability estimation from early failure data I. *Proceedings of the Fourth International Conference on Modelling and Simulation*, 386-390.

Summit, R.A. Cerone, P. and Diamond, N.T. (2002b), Simulating reliability estimation from early failure data II. *Proceedings of the Fourth International Conference on Modelling and Simulation*, 391-396.

Suzuki, K. (1985), Estimation of lifetime parameters from incomplete field data. *Technometrics*, **27**, 263-271.

Thomas, M.U. (1999), Some economic decision problems in warranty planning. *The Engineering Economist*, **44**, 184-196.

Tijms, H.C. (1994), *Stochastic Models: An Algorithmic Approach*. John Wiley and Sons, New York.

Trivedi, K.S. (1982), *Probability and Statistics with Reliability, Queuing and Computer Science Applications*, Prentice-Hall, Englewood Cliffs, New Jersey.

Tukey, J.W. (1977), *Exploratory Data Analysis*. Addison-Wesley, Reading, Massachusetts.

Venables,W.N. and Ripley, B.D. (1999), *Modern Applied Statistics with S-Plus*. Springer-Verlag, New York.

Wang, C.H. and Sheu, S.H. (2003), Optimal lot sizing for products sold under free-repair warranty. *European Journal of Operational Research*, **149**, 131-141.

Wasserman,G.S. (1992), An application of dynamic linear models for predicting warranty claims. *Computers and Industrial Engineering*, **22**, 37-47.

Wasserman, G.S. and Sudjianto, A. (1996), A comparison of three strategies for forecasting warranty claims. *IIE Transactions*, **28**, 967-977.

White, J.S. (1964), Weibull renewal analysis. *Proceedings of the Aerospace Reliability and Maintainability Conference*, New York, Society of Automotive Engineers, 639-657.

Wonnacott, T.H. and Wonnacott, R.J. (1990), *Introductory Statistics* (Fifth Edition). John Wiley and Sons, New York.

Wu, H. and Meeker, W.Q. (2002), Early detection of reliability problems using information from warranty databases. *Technometrics*, **44**, 120-133.

Yohai, V., Stahel, W.A. and Zamar, R.H. (1991), A procedure for robust estimation and inference in linear regression, *Directions in Robust Statistics and Diagnostics, Part II*, W.A. Stahel and S. Weisberg (Editors). Springer-Verlag.

Xie, M. (1989), On the Solution of Renewal-Type Integral Equations. *Communications in Statistics*, **18**, 281-293.

Appendix A: S-Plus Scripts for Chapter 4

This appendix contains scripts of functions, and a function's output, that are referred to in Chapter 4. They are included here for completeness and ready reference. Other scripts and datasets that are referred to in Chapter 4 that are not included here can be found on the accompanying compact disc.

A.1 Exploratory Data Plots and Summaries (*PlotSummary.ssc*)

```
#   FUNCTION TO PLOT AND SUMMARISE DATA IN A DATA FRAME

#------------
# DESCRIPTION
#------------
# OVERVIEW
# ========
# This function is used for exploratory data analysis. It will produce:
# (1) several plots, including boxplots and scatterplots;
# (2) a statistical summary of the data in the data frame.

# INPUT
# =====
# A data frame for which plots and summaries are required.

# -------
# PROGRAM
# -------
"fPlotSmr"<-
function(ClaimsDF)
{
# PLOTS
# =====
# Boxplots of numeric data.
   boxplot( dates(ClaimsDF$prod,origin=c(1,1,1997)), dates(ClaimsDF$sale,origin=c(1,1,1997)),
      dates(ClaimsDF$repr,origin=c(1,1,1997)), names = c("prod", "sale", "repr"),
      xlab = "Date", ylab = "Days since 1/1/97", style.bxp="old")
   title("Dates Boxplots")

   boxplot(ClaimsDF$dist/1000, names = "dist", ylab = "x1000km", style.bxp="old")
   title("Distance Boxplot")

   boxplot(ClaimsDF$age, names = "age", ylab = "Months", style.bxp="old")
   title("Age Boxplot")

   boxplot(ClaimsDF$age2, names = "age2", ylab = "Days", style.bxp="old")
   title("Age2 Boxplot")

   boxplot(ClaimsDF$cost, names = "cost", ylab = "$", style.bxp="old")
   title("Cost Boxplot")

# Matrix plot of Date Fields.
   pairs(cbind(ClaimsDF$prod, ClaimsDF$sale, ClaimsDF$repr), labels=c("Prod","Sale","Repair"))

# Distance vs Age plot.
   plot(ClaimsDF$age, ClaimsDF$dist/1000, xlab = "age (months)", ylab = "dist (1000km)")
   title("Distance vs Age Plot")

# Distance vs age2 plot.
   plot(ClaimsDF$age2, ClaimsDF$dist/1000, xlab = "age2 (days)", ylab = "dist (1000km)")
   title("Distance vs age2 Plot")

# OUTPUT
# ======
# A list of summary statistics, and a linear regression model through the origin for dist vs
# age are provided.
   list(summary(ClaimsDF),
   lm(dist ~ age+(-1), data = ClaimsDF, na.action = na.exclude))
```

A.2 Function to Check the Data (*Check.ssc*)

```
#                         FUNCTION TO CHECK DATA

#------------
# DESCRIPTION
#------------
# PURPOSE
# =======
# To identify records that contain an error in the dates, age2, age and dist fields.

# OVERVIEW
# ========
# The input of this function is a claims data frame.
# The output is a report of possible errors in the data frame.
# This function should be followed by the cleaning function fClean.

# -------
# PROGRAM
# -------
"fCheck"<-
function(ClaimsDF)
{  attach(ClaimsDF) # Enables variables in dataframe to be accessed directly.

# This function assumes values in VIN for all cases.
   if(any(is.na(VIN)))
      stop(message = "This function will not work properly if VIN contains any NAs")

# Convert date columns to numeric to remove date attributes from output.
   prod <- as.numeric(prod)
   sale <- as.numeric(sale)
   repr <- as.numeric(repr)

# SET UP LINEAR MODELS OF dist BASED ON age AND age2
# =================================================
# This establishes a linear model for dist based on age, to be used in a later section.
   model <- lm(dist ~ age+(-1), na.action=na.exclude)
   rate <- model$coefficients
   num <- length(VIN[!is.na(age) & !is.na(dist)])
   s <- sqrt( sum(model$residuals^2)/(num - 1) )
   sumSqrd <- sum(age^2, na.rm=T)

# This establishes a linear model for dist based on age2, to be used in a later section.
   model2 <- lm(dist ~ age2+(-1), na.action=na.exclude)
   rate2 <- model2$coefficients
   num2 <- length(VIN[!is.na(age2) & !is.na(dist)])
   s2 <- sqrt( sum(model2$residuals^2)/(num2 - 1) )
   sumSqrd2 <- sum(age2^2, na.rm=T)

# This list will output required statistics
   result <- list(

# IDENTIFY DATE FIELDS WITH NAs
# =============================
   "Production date is NA in these cases" = row.names(ClaimsDF[is.na(prod), ]),
   "Sale date is NA in these cases" = row.names(ClaimsDF[is.na(sale), ]),
   "Repair date is NA in these cases" = row.names(ClaimsDF[is.na(repr), ]),

# IDENTITY DATE FIELDS WITH 0
# ===========================
   "Production date is 0 in these cases" = row.names(ClaimsDF[prod==0 & !is.na(prod), ]),
   "Sale date is 0 in these cases" = row.names(ClaimsDF[sale==0 & !is.na(sale), ]),
   "Repair date is 0 in these cases" = row.names(ClaimsDF[repr==0 & !is.na(repr), ]),

# IDENTITY CASES WHERE SALE OR REPAIR IS BEFORE FIRST PRODUCTION DATE
# ===================================================================
   "Sale is before first production date in these cases"
      = row.names(ClaimsDF[sale < min(prod, rm.na=T) & !is.na(sale), ]),
```

```
   "Repair is before first production date in these cases"
      = row.names(ClaimsDF[repr < min(prod, rm.na=T) & !is.na(repr), ]),

# IDENTITY RECORDS WHERE SALE OR REPAIR IS BEFORE PRODUCTION
# ===========================================================
   "Sale is before production in these cases"
      = row.names(ClaimsDF[sale < prod & !is.na(sale) & !is.na(prod), ]),
   "Repair is before production in these cases"
      = row.names(ClaimsDF[repr < prod & !is.na(repr) & !is.na(prod), ]),

# IDENTITY CASES WHERE PRODUCTION < REPAIR < SALE
# ================================================
   "The number of cases where production < repair < sale is"
      = length(VIN[repr<sale & repr>prod & !is.na(repr) & !is.na(sale)]),
   "These may be presale, dealer-initiated claims, where age = 0. Of those, age > 0 in case:"
      = row.names(ClaimsDF[repr<sale & repr>prod    & !is.na(repr) & !is.na(sale) & age > 0
      & !is.na(age),]),
   "Also, distance should be small. Of those, distance > 100 in case"
      =row.names(ClaimsDF[repr<sale & repr>prod & !is.na(repr) & !is.na(sale) & dist > 100
      & !is.na(dist),]),

# IDENTIFY RECORDS WHERE SALE IS TOO LONG AFTER PRODUCTION
# =========================================================
# There is a lag between production and sales. Records with a lag greater than 1 year may be in
# error. However, if age2 matches age, then the sale date is probably correct, and there was
# just a long lag. Exclude cases where sale or prod are NA.
   "The number of cases with sale delay (=sale-prod) > 1 year is"
      =length(VIN[sale>prod+365 & !is.na(sale) & !is.na(prod)]),

# IDENTIFY CASES WHERE age2 DIFFERS FROM age BY > 1 MONTH
# ========================================================
   "Number of cases where age2 differs from age by > 1 month & repair is after sale is"
      = length(VIN[abs(age2 - age*30.4) > 31 & !is.na(age2) & age2 >= 0]),
   "The cases are"
      = row.names(ClaimsDF[abs(age2-age*30.4) > 31 & !is.na(age2) & age2 >= 0 & age == 0,]),

# IDENTIFY CASES WHERE age2 DIFFERS FROM age BY > 2 MONTH
# ========================================================
   "Number of cases where age2 differs from age by > 2 months & repair is after sale is"
      = length(VIN[abs(age2 - age*30.4) > 61 & !is.na(age2) & age2 >= 0]),
   "The cases are"
      = row.names(ClaimsDF[abs(age2 - age*30.4)>61 & !is.na(age2) & age2>=0 & age==0,]),

# CHECK age AND age2 AGAINST dist
# ================================
# Using the linear models of dist based on age and age2 established above, the value for dist
# is checked against the 99% prediction interval for the values of age and age2.
   "Number of cases where distance is outside the 99% prediction interval based on age is"
      =length(VIN[abs(dist - rate*age) >
          qt(0.995, num - 2) * s * sqrt(1 + age^2/sumSqrd) &
          !is.na(age) & !is.na(dist)] ),
   "Number of cases where distance is outside the 99% prediction interval based on age2 is"
      =length( VIN[abs(dist - rate2*age2) >
          qt(0.995, num2 - 2) * s2 * sqrt(1 + age2^2/sumSqrd2) &
          !is.na(age2) & !is.na(dist)] ),

# OBTAIN COSTS ABOVE $20,000
# ==========================
   "The repair costs above $20,000 are"
      =cost[cost > 20000]
   )

# CLOSING ROUTINE
# ===============
   detach("ClaimsDF")
# The next line returns the above list.
   result
}
```

A.3 Output of *fCheck*

```
> fCheck(claims97)
$"Production date is NA in these cases":
character(0)

$"Sale date is NA in these cases":
[1] "413"   "414"   "1613"  "2456"  "2943"  "11423" "12631" "39537"

$"Repair date is NA in these cases":
character(0)

$"Production date is 0 in these cases":
character(0)

$"Sale date is 0 in these cases":
character(0)

$"Repair date is 0 in these cases":
character(0)

$"Sale is before first production date in these cases":
character(0)

$"Repair is before first production date in these cases":
character(0)

$"Sale is before production in these cases":
[1] "4525"  "4526"  "10824" "10825" "23679"

$"Repair is before production in these cases":
 [1] "8240"  "12753" "13897" "18411" "19559" "25744" "27964" "27987" "28667" "31983"
[11] "33201" "33203" "43473" "50289" "52903" "58952" "61433"

$"The number of cases where production < repair < sale is":
[1] 753

$"These may be presale, dealer-initiated claims, where age = 0. Of those, age > 0 in
case:":
[1] "2067"  "34210" "39868" "60277"

$"Also, distance should be small. Of those, distance > 100 in case":
 [1] "2067"  "5579"  "7763"  "7843"  "7844"  "8108"  "10767" "10768" "10770" "10775"
[11] "13689" "14360" "25903" "28477" "28550" "30633" "30762" "35262" "35297" "35461"
[21] "35640" "38504" "39865" "39868" "39923" "41298" "41855" "43121" "44958" "47312"
[31] "58302" "58553" "59475" "60278" "61804" "62345"

$"The number of cases with sale delay (=sale-prod) > 1 year is":
[1] 120

$"Number of cases where age2 differs from age by > 1 month & repair is after sale is"
:
[1] 111

$"The cases are":
 [1] "2192"  "10825" "12135" "23258" "24872" "24873" "24874" "25743" "28704" "32987"
[11] "32988" "32989" "39447" "39448" "49117" "53164"

$"Number of cases where age2 differs from age by > 2 months & repair is after sale is
":
[1] 77

$"The cases are":
 [1] "2192"  "10825" "12135" "23258" "24872" "24873" "24874" "25743" "28704" "32987"
[11] "32988" "32989" "39447" "39448" "49117" "53164"

$"Number of cases where distance is outside the 99% prediction interval based on age
is":
```

```
[1] 585

$"Number of cases where distance is outside the 99% prediction interval based on age2
 is":
[1] 715

$"The repair costs above $20,000 are":
   32166 53847
 23636.3 27193
```

A.4 Function to Clean the Data (*Clean.ssc*)

```
#                       FUNCTION TO CLEAN DATA

#------------
# DESCRIPTION
#------------
# OVERVIEW
# ========
# This function is used to clean the data in a data frame. It should follow the check function
# fCheck. It changes incorrect values in age2 and puts a flag indicating the change in the
# "altered" field. Suspected errors where age2 has not been changed are flagged in the
# "error" field.

# OUTLINE
# =======
# 1. Check for errors in repair date.
# 2. Check for errors in sale.
# 3. Check if age2 < 0.
# 4. Compare age2 with age or distance.

# -------
# PROGRAM
# -------
"fClean"<-
function(ClaimsDF)
{
   attach(ClaimsDF)

# ESTABLISH CONDITIONS FOR AGE AND DIST TO BE WITHIN OUTLIER LIMITS
# ================================================================
# Use a non-plot box function to obtain the statistics for outliers
   ageBox <- boxplot(age, plot=F)
   distBox <- boxplot(dist, plot=F)

# Set conditions for age and dist to be less than outliers. In many records where age=0, the
# values in dist or age2 conflict with this value, so do not include these amongst the "good"
# records containing non-outliers. Note: ageBox$out[1] is the smallest age outlier.
   ageGood <- age !=0 & age < ageBox$out[1]    & !is.na(age)
   distGood <- dist < distBox$out[1] & !is.na(dist)

# OBTAIN USAGE RATE (ie. dist = rate2 x age2)
# ========================================
# Robust model used to remove the outlier effects.
   model2 <- lmRobMM(formula = dist~age2+(-1), data = ClaimsDF, na.action=na.exclude)
   rate2 <- model2$coefficients

# CHECK FOR ERRORS IN REPAIR DATE
# ==============================
# Age2 has been calculated as repr - sale. Hence an error in either of these two fields will
# result in an error in age2. Errors in repr are cleaned in this section, whilst errors in
# sale are cleaned in the next section. Records where prod < repr < sale are looked at later.

# If repr is NA, or if repr < prod (this covers repr=0), age2 will be in error.
   reprBad <- is.na(repr) | repr < prod

# Age2 will be given a value as follows:
#1. If age is not 0, 999 or an outlier, then age2 gets its value from age.
```

```
   reprBadAgeGood <- reprBad & ageGood
   ClaimsDF$age2[reprBadAgeGood] <- round(age*365.25/12, digit=0)[reprBadAgeGood]
   ClaimsDF$altered[reprBadAgeGood] <- "reprErr.AgeUsed"

#2. Else if dist is not an outlier, then age2 gets its value from dist.
   reprBadDistGoodAgeBad <- reprBad & distGood & !ageGood
   ClaimsDF$age2[reprBadDistGoodAgeBad] <- round(dist/rate2, digit=0)[reprBadDistGoodAgeBad]
   ClaimsDF$altered[reprBadDistGoodAgeBad] <- "reprErr.DistUsed"

#3. Else age2 gets NA.
   reprBadAgeBadDistBad <- reprBad & !ageGood & !distGood
   ClaimsDF$age2[reprBadAgeBadDistBad] <- NA
   ClaimsDF$altered[reprBadAgeBadDistBad] <- "reprErr.NAused"

# CHECK FOR ERRORS IN SALE DATE
# ============================
# Define delay as the difference between sale and prod. Look for outliers in delay. Because it
# is skewly distributed, take log(delay) firstly. Remove cases where sale is NA or < prod.
   delay <- log( (sale - prod + 1)[!is.na(sale) & sale >= prod] )
      # + 1 ensures that we do not get log(0)
   delayBox <- boxplot(delay,plot=F)
   longDelay <- exp(delayBox$stats[1])     #This is the largest delay that is not an outlier
   medDelay <- round(exp(delayBox$stats[3]), digit=0)

# Sale will be in error if it is NA, < prod (this covers 0) or if sale is too long after prod.
# Exclude records where age2 was given a new value above because repr was in error.
   saleBad <- (is.na(sale) | sale < prod | sale > prod + longDelay) & !reprBad

# When sale is in error (and repr is okay), age2 will be given a value as follows:
#1. If dist is not an outlier, age2 gets its value from dist. Dist gives the largest number of
#   correlations of two fields amongst age, dist and repr - (prod + medDelay).
   saleBadDistGood <- saleBad & distGood
   ClaimsDF$age2[saleBadDistGood] <- round(dist/rate2, digit=0)[saleBadDistGood]
   ClaimsDF$altered[saleBadDistGood] <- "saleErr.DistUsed"

#2. If this new value of age2 doesn't match either age or repr - (prod + medDelay), and age
#   matches repr - (prod + medDelay), age2 gets repr - (prod + medDelay).
   saleBadDelayMatchAge <- saleBadDistGood & (abs(age*365.25/12-ClaimsDF$age2)>31 |
      abs(repr-(prod+medDelay)-ClaimsDF$age2)>31) &
      abs(repr-(prod+medDelay)-age*365.25)<32
   ClaimsDF$age2[saleBadDelayMatchAge] <-pmax(0,repr-(prod+medDelay))[saleBadDelayMatchAge]
   # If repr < (prod+medDelay), age2 gets 0.
   ClaimsDF$altered[saleBadDelayMatchAge] <- "saleErr.DelayUsed"

#3. If dist is an outlier or NA, age2 gets repr - (prod + medDelay).
   saleBadDistBad <- saleBad & !distGood
   ClaimsDF$age2[saleBadDistBad] <- pmax(0, repr - (prod + medDelay))[saleBadDistBad]
   ClaimsDF$altered[saleBadDistBad] <- "saleErr.DelayUsed"

# CHECK FOR CASES WHERE REPAIR OCCURS BETWEEN PRODUCTION AND SALE
# ===============================================================
# If prod < repr < sale, and age2 has not already been changed above, age2 gets 0.
   reprB4sale <- repr < sale & repr >= prod & ClaimsDF$altered == ""
   ClaimsDF$age2[reprB4sale] <- 0
   ClaimsDF$altered[reprB4sale] <- "reprB4sale.0used"

# LOOK FOR CORRELATION BETWEEN AGE2, AGE AND DIST
# ==============================================
# This is needed to obtain a 99% prediction interval for dist.
   num2 <- length(ClaimsDF$VIN[!is.na(age2) & !is.na(dist)])
   s2 <- sqrt(sum(model2$residuals^2) / (num2-1) )
   sumSqrd2 <- sum(ClaimsDF$age2^2, na.rm=T)

# Create a model for dist based on age.
   model <- lmRobMM(dist~age+(-1), data = ClaimsDF, na.action=na.exclude)
   rate <- model$coefficients
   num <- length(ClaimsDF$VIN[!is.na(age) & !is.na(dist)])
   s <- sqrt(sum(model$residuals^2) / (num-1) )
```

```
   sumAgeSqrd <- sum(ClaimsDF$age^2, na.rm=T)

# If age2 does not match either age or dist, check if age matches dist. In this case put age
# in age2. Allow 32 days difference between age and age2 to cover several cases that are just
# over one month.
   ageDistMatch <- abs(ClaimsDF$age * 365.25 / 12 - ClaimsDF$age2) > 32 &
      abs(ClaimsDF$dist - rate2*ClaimsDF$age2)   >
         qt(0.995, num2-2) * s * sqrt(1+ClaimsDF$age2^2 / sumSqrd2) &
      abs(ClaimsDF$dist - rate * ClaimsDF$age)   <=
         qt(0.995, num-2) * s * sqrt(1+ClaimsDF$age^2 / sumAgeSqrd)

   ClaimsDF$age2[ageDistMatch] <- round(age*365.25/12, digit=0)[ageDistMatch]
   ClaimsDF$altered[ageDistMatch] <- "age2~DistErr.ageUsed"


# CHECK COST FOR ERRORS
# ====================
# Flag cases where the cost < 0 or > $10,000. Cost outliers were not used because there were
# a number of them.
   costBad <- cost < 0  | cost > 10000
   ClaimsDF$error[costBad] <- "costExtreme"


# CLOSING ROUTINE
# ===============
   detach("ClaimsDF")
   ClaimsDF
}
```

A.5 Function to Add Sales Records (*AddSale.ssc*)

```
#  CHECK IF ANY VEHICLES IN CLAIMS DATABASE NOT LISTED IN SALES DATABASE

# PURPOSE
# =======
# This function adds records to the sales database from records listed in the claims database.

# PROGRAM
# =======
# A logical vector that identifies which records in claims97 are not in sales97.
notInSales <- !is.element(claims97$VIN, sales97$VIN)

if( length(table(notInSales)) > 1 ) # This occurs only if there are Ts and Fs in the table.
{  addSales <- cbind(claims97[notInSales, c("VIN","prod")], region="",prodn="-1",age2=1096)
   guiModify( "double", Name = "addSales$age2",  Precision = "0")
   dups <- duplicated(addSales$VIN)
   addSales <- addSales[!dups,]

# Combine the original sales97 data frame with the records in claims not in sales.
sales97 <- rbind(sales97, addSales)
}

# CLEAN UP
# ========
# First check that sales97ORIG has been successfully written to the set directory.
rm(notInSales, addSales, dups)
```

A.6 Function to Calculate Components' Ages (*CompAge.ssc*)

```
#     CALCULATION OF age2 & dist2 IN VEHICLES WITH REPEAT FAILURE OF A COMPONENT

#------------
# DESCRIPTION
#------------
# OVERVIEW
# ========
# This function calculates the age and dist of components fitted to a vehicle for a second or
# subsequent time. It assumes that the data frame is sorted by VIN, then by part, then by repr.
```

```
# A data frame with an updated age2 and added dist2, the distance the component has travelled,
# is returned.

# Note: The attach function makes copies of the fields of the data frame in the working
# directory. To make changes in the data frame, you need to prefix the field names with the
# data frame name. Likewise, any changes in the dataframe fields are not made in the "copied"
# objects.

# INPUT
# =====
# 1. CleanDF = a data frame of cleaned data. Default = clean97.
# 2. Wrnty = warranty length in years. Default = 3.

# OUTPUT
# ======
# A data frame with component ages calculated.

# -------
# PROGRAM
# -------
"fCompAge"<-
function(CleanDF= clean97, Wrnty=3)
{
   attach(CleanDF)

# Convert warranty to days.
   Wrnty <- round(Wrnty * 365.25)

# CALCULATE DIFFERENCE IN age2 IN MULTIPLE REPAIR CASES
# ====================================================
# The following identifies records that have the same VIN as the previous record.
   sameVIN <- duplicated(VIN)

# The following identifies records that have the same part as the previous record. Note: the
# function duplicated() is not suitable for use with part, because it will detect any previous
# occurrences of the particular part, not just in the last record
   samePart <- vector(mode="logical", length=length(VIN))
   n <- length(VIN)
   samePart[2:n] <- part[1:(n-1)]==part[2:n]

# This identifies records that have the same VIN and part as the previous record, ie, records
# where a component has been replaced a second or subsequent time. It excludes records where
# part is blank or NA.
   sameVINpart <- sameVIN & samePart & part!="" & !is.na(part)

# To put the age of a component into age2, we need the differences in age2 between subsequent
# records. This will have a length one less than the length of age2 itself, so add a dummy
# record at the start, as this cannot be a subsequent repair.
   diffAge2 <- c(0, diff(age2))
# Assign these differences in age2 to the records where sameVINpart is true.
   CleanDF$age2[sameVINpart] <- diffAge2[sameVINpart]

# Create field dist2 for the usage of a component. Set it to dist initially, then calculate
# difference from previous repair in mulltiple repair vehicles.
   dist2 <- dist
   diffDist <- c(0, diff(dist))
   dist2[sameVINpart] <- diffDist[sameVINpart]

# There are errors in some of the data, where the second repair on a vehicle has a recorded
# distance that is less than the previous record (ie, the last repair). As this cannot be the
# case, we will change these to NAs.
   dist2[dist2<0] <- NA

# If there are any records with 0 distance, make them 1 so that we can take log of them.
   dist2[dist2==0] <- 1

# ADD AND FORMAT FIELDS
# ====================
```

```
# Add this field to the data frame.
   CleanDF <- cbind(CleanDF, dist2)

# ADJUST compAge TO EXCLUDE POST-WARRANTY CLAIMS
# ==============================================
# Some claims have been made after 3 years. These will be excluded becuase most vehicles'
# repairs are unknown after 3 years. Using the post-warranty claim data introduces a bias.
# Note that the analysis is based on age2 and not on distance, so no check is made on distance.
   CleanDF <- CleanDF[CleanDF$age2 <= Wrnty, ]

# CLOSING ROUTINE
# ===============
   detach("CleanDF")
   CleanDF
}
```

A.7 Function to Construct a Survival Database (*Survive.ssc*)

```
#      FUNCTION TO COMBINE CLAIMS AND SALES DATA FRAMES FOR SURVIVAL ANALYSIS

#------------
# DESCRIPTION
#------------
# OVERVIEW
# ========
# This function is to be called by individual part (part numbers). This can be done by indexing
# the data frame on a part# (ie. CompDF[CompDF$part=="part#",]) or by using the S-Plus
# function by().

# This function assumes that CompDF is sorted by VIN, and will not work properly unless this
# is the case. This function is to be run after cleaning the claims data with fClean, and
# obtaining the age of components in multi-repaired vehicles with fCompAge.

# INPUT
# =====
# 1. CompDF = a data frame containing data for one component.
# 2. SalesDF = a data frame of vehicles produced. Default = sales97.
# 3. Wrnty = the warranty length in years. Default = 3.

# OUTPUT
# ======
# This function returns a data frame of all vehicles and components and includes a status
# field of failed (=1) or censored (=0), ready for survival analysis.

# -------
# PROGRAM
# -------
"fSurvive"<-
function(CompDF, SalesDF=sales97, Wrnty=3)
{
# CHECK INPUT PARAMETERS
# ======================
   if(length(CompDF$VIN) == 0)
      stop(message="No components with that part number.")
   if(length(table(CompDF$part)) != 1)
      stop(message="This function requires the input CompDF to have only one part.")
   if(length(SalesDF$VIN) == 0)
      stop(message="The inputted sales database is empty.")
   if( all.equal(order(CompDF$VIN), 1:length(CompDF$VIN)) ) T
      else stop(message="This function requires the inputted database to be sorted by VIN.")

# Convert warranty to days
   Wrnty <- round(Wrnty * 365.25)

# DATA FRAME OF FAILED COMPONENTS FROM CLAIMS DATA FRAME
# ======================================================
# Each entry in the claims database represents a failed component, which is then written as a
# failure record in the survival database. Pick up the fields VIN, age2, dist2, part, partname
```

```
# and add a status of 1.
   Failed   <- cbind(CompDF[ ,c("VIN", "age2","dist2","part","partname","cost")], status=1)

# Some of these failures occur after 3 years. If left in, these points result in a high
# failure rate being evaluated beyond 3 years because the bulk of the data is censored by
# then. Thus, to avoid this bias in the model, it is necessary to censor these failures at 3
# years. The status needs to be changed to 0 for censored also. Note that fCompAge has already
# restricted age2 to the warranty length, so don't need to do it here, but it has been left in.
   beyond3yrs <- Failed$age2 > Wrnty
   Failed$age2[beyond3yrs] <- Wrnty
   Failed$dist2[beyond3yrs] <- Failed$dist[beyond3yrs] / Failed$age2[beyond3yrs] * Wrnty
   Failed$status[beyond3yrs] <- 0

# DATA FRAME OF CENSORED COMPONENTS FROM CLAIMS DATA FRAME
# ========================================================
# Each replaced component in the claims database that does not fail again needs to be written
# as a censored item in the survival database.

# The following identifies records that have the same VIN as the previous record. These are
# vehicles that have had more than one repair on a particular part. Note that only one part
# should have been passed to this function, so duplicate VINs identify subsequent repairs.
   sameVIN <- duplicated(CompDF$VIN)

# Components that are subsequently replaced will have the same VIN as the next, rather than the
# previous record. These cases can be identified by moving sameVIN up by one record. The last
# record will not a subsequent repair, so it is given an F. However, if there is only one
record
# there is no need to do anything. (The next line works only if there are at least 2 records.)
   if(length(sameVIN) > 1)
       sameVIN<- c(sameVIN[2:length(sameVIN)], F)

# The censored items, where the replaced component has not yet failed, are the ones where
# sameVIN is F. Pick up only the fields that are required for survival analysis.
   Censored <- CompDF[!sameVIN,c("VIN","sale","repr","age2","dist2","part","partname")]

# Calculate max age of component whilst vehicle is still under warranty. Set componentAge to
the
# age of the component, for later use.
   componentAge <- Censored$age2

# Set age2 to difference between warranty end date (= sales date + 3 years) and repair date, as
# component may get repaired elsewhere after warranty ends. Note:
# 1. pmax() is used because repair date may be after warranty ends, which means that
#    (Censored$sale + Wrnty - Censored$repr) ends up negative.
# 2. pmin() is used in case the warranty end date is afer 21/5/2002, which was when the
database
#    was read.
   Censored$age2 <- pmin(pmax(0,as.numeric(Censored$sale + Wrnty - Censored$repr)),
       as.numeric(dates("21/5/2002",format="d/m/y")-Censored$repr))

# Set dist2 to usage rate * age, with a limit of 100,000km.
   Censored$dist2 <- pmin(100000, round((Censored$dist2/componentAge)*Censored$age2),
digits=0)

# Remove sale and repr, as they are no longer needed and add status = 0 and cost = NA.
   Censored<-cbind(Censored[,c("VIN", "age2","dist2","part","partname")], status=0, cost=NA)

# CREATE A DATA FRAME OF CENSORED DATA FROM SALES DATA FRAME
# ==========================================================
# Some vehicles have had no components fail. These can be obtained by finding VINs in SalesDF
# that are not in CompDF. These are written into the survival database as censored items
# with part = none. The age of these components should be the warranty length if the sales data
# has been collected after the warranty has run out on all vehicles. If not, the sales data
# contains an estimate of the vehicles' ages. The age of vehicles with claims can be obtained
# from compAge using: "21/5/2002" (download date) - (repr - age2), where (repr - age2) is equal
# to sale if sale has been correctly entered.
   NoClaims <- SalesDF[!is.element(SalesDF$VIN,unique(CompDF$VIN)),c("VIN","age2")]
```

```
# Add dist2 = NA, since claims data will have this field, and add status = 0.
   NoClaims <- cbind(NoClaims, dist2=NA, part="", partname="", status=0, cost=NA)

# COMBINE THE THREE DATA FRAMES FOR SURVIVAL ANALYSIS
# ===================================================
   CombinedDF <- rbind(Failed, Censored, NoClaims)

# Convert status to integer type.
CombinedDF <- convert.col.type(target = CombinedDF,
   column.spec = "status",
      column.type = "integer" )

# Return CombinedDF ready for survival analysis.
CombinedDF
}
```

Appendix B: S-Plus Scripts for Chapter 5

This appendix contains the script of function that has been used to establish the usage pattern of vehicles. The script is referred to in Chapter 5.

B.1 Calculation of Usage Rates (*Usage.ssc*)

```
#      VEHICLE USAGE RATES

#------------
# DESCRIPTION
#------------
# PURPOSE
# =======
# This function calculates the usage rates of vehicles in Claims and obtains limits for
# usage that are not outliers.

# INPUT
# =====
# Claims = the claims data (default = claims97)
# ageLmt = the minimum age for usage rate calculation. (Default = 61.) Lower values than this
# are excluded to increase accuracey of usage rate calculation
# distLmt the minimum distance for usage rate calculation. (Default = 3300.) Again, this is
# done to get a reasonable usage rate figure.

# OUTPUT
# ======
# 1. A boxplot and a density plot of vehicle usages.
# 2. A list containing a table indicating the number of usable age2, dist and combined values,
#  and the  maximum non-outlier value of usage.

# -------
# PROGRAM
# -------
"fUsage" <-
function(Claims=claims97, ageLmt=61, distLmt=3300)
{
# Select only the columns needed to calculate usage.
   Claims <- Claims[ , c("VIN", "age2", "dist")]

# Claims may contain vehicles with more than one claim. To pick up a vehicle once only, sort
# Claims by VIN and age2 in descending order. This will give the biggest age2 first for a
# given, which will be the one used to calculate usage.
   Claims <- sort.col(target = Claims,  columns.to.sort = "@ALL",
       columns.to.sort.by = c("VIN", "age2"), ascending = F)

# Records containing a vehicle's second or subsequent claim will be identified by a "T".
   dups <- duplicated(Claims$VIN)

# Only use non-duplicated records to work out a vehicle's usage.
   Claims <- Claims[!dups,]

# Select values of age2 that are not not outliers. As a number of claims are made before sale
# of a vehicle, exclude values of age2 < ageLmt (default = 2 months), as these values will not
# give accurate usage figures.
   age2Box <- boxplot(Claims$age2, plot=F )
   goodAge <- Claims$age2 <= age2Box$stats[1] & Claims$age2 >= age2Box$stats[5] &
       Claims$age2 > ageLmt & !is.na(Claims$age2)

# Similarly, select data where dist is not an outlier and is over distLmt, which is the
# distance travelled in two months by a vehicle with an average usage rate of 19,700km/yr.
   distBox <- boxplot(Claims$dist, plot=F )
   goodDist <- Claims$dist <= distBox$stats[1] & Claims$dist >= distBox$stats[5] &
       Claims$dist > distLmt & !is.na(Claims$dist)

# Select only the data where both age2 and dist have good values to calculate usage.
```

```
   goodData <- goodAge & goodDist
   usage <- 365.25*Claims$dist[goodData]/Claims$age2[goodData]
   usageBox <- boxplot(usage, plot=F)

# Output.
   boxplot(usage, style.bxp = "old",ylab="Usage (km/yr)")
   title("Vehicle Usage Rates")

   plot(Claims$age2[goodData], Claims$dist[goodData]/1000,
      xlab="Age (days)", ylab="Distance (x1000km)", page="new")
   lines(c(1096,1096,0), c(0,100,100), col=8)
   lines(c(0,1.2*1096), c(0,1.2*100), col=6)
   title("Vehicle Usage Rates")

   numUsage <- length(usage)
   list(
      "Table of usable age2" = table(goodAge),
      "Table of usable dist" = table(goodDist),
      "Table of usable data" = table(goodData),
      "Max non-outlier usage" = usageBox$stats[1],
      "Proportion of users above this value" = length((1:numUsage)[usage > 100000/3])/numUsage
      )
}
```

Appendix C: S-Plus Scripts for Chapter 6

This appendix contains scripts of functions that have been used in establishing the parameters of the warranty cost models, and are referred to in Chapter 6. They are included here for completeness and ready reference. Other scripts and datasets that are referred to in Chapter 6 that are not included here can be found on the accompanying compact disc.

C.1 Exponential Modelling (*CostExp.ssc*)

```
#      POINT & INTERVAL WARRANTY COST ESTIMATES USING AN EXPONENTIAL MODEL

#------------
# DESCRIPTION
#------------
# PURPOSE
# =======
# The exponential distribution is used to model failure of components with claims frequency
# between ClaimsFrom and ClaimsTo during the warranty period. An estimate of the warranty cost
# and its variance are made.

# OUTLINE
# =======
# This program is in four parts:
# 1. The number of components, the sum of the average costs of repair of these components and
#   an upper limit on these sums are obtained.
# 2. The failure rate and an upper limit for it are obtained. These statistics depend on the
#   part chosen because there will be variations on the timing of failures.
# 3. The expected warranty cost and an upper limit for it are obtained.

# INPUT
# =====
# 1. The data frame is passed in as the parameter "CompDF". (Default = compAge.)
# 2. ClaimsFrom = The minimum number of claims to be modelled by the exponential distribution.
#   (Default = 1.)
# 3. ClaimsTo = The maximum number of claims to be modelled by the exponential distribution.
#   Should be no more than 20 (the default).
# 4. warranty = The warranty length in years. (Default = 3.)
# 5. Production = The number of vehicles produced in a year. (Default = 30138.)

# OUTUT
# =====
# The output is a dataframe containing the following:
# Part details: part id (Part), part name (Name), number of claims (Claims);
# Repair cost details: cost of component replacement (CmpCst) and its SD (CmpCstSD);
# Expoential model details: the failure rate (Rate) and its SD (RateSD), the expected number of
#   repairs during the warranty period (NumRep), and the expected warranty cost (WrntCst)
#   and its SD (WrntCstSD) and variance (WrntCstExp);

# -------
# PROGRAM
# -------
"fCstExp"<-
function(CompDF=compAge, SalesDF=sales97, ClaimsFrom=1, ClaimsTo=20, Wrnty=3, Production=30138)
{

# CHECK INPUT PARAMETERS
# =====================
    if(length(CompDF$VIN) == 0)
        stop(message="No components with that part number.")
    if(length(SalesDF$VIN) == 0)
        stop(message="The inputted sales database is empty.")
    if(Wrnty <= 0)
        stop(message="The parameter 'Wrnty' (warranty period in years) should be positive.")
    if(Production <= 0 )
        stop(message="The parameter 'Production' should be positive.")
    if(ClaimsFrom > ClaimsTo)
```

```
      stop(message="The parameter 'ClaimsFrom' cannot be greater than 'ClaimsTo'.")
   if( ClaimsFrom != round(ClaimsFrom,0) | ClaimsTo != round(ClaimsTo,0) )
      stop(message="The parameters 'ClaimsFrom' and 'ClaimsTo' must be positive integers.")
   if(ClaimsTo > 20)
      warning("Components have more than 20 claims. Exponential model may not be suitable")

# CREATE VARIABLES AND PERFORM CHECKS
# =================================
# If table of parts doesn't already exist create it .
   if( !exists("partFrq") )
      partFrq <- sort( table(CompDF$part), na.last=F )

# Create data frame to hold the data.
   output <- data.frame(Part=NA, Name=NA, Claims=NA, CmpCst=NA,CmpCstSD=NA,
               Rate=NA, RateSD=NA, NumRep=NA, WrntCst=NA, WrntSD=NA, WrntVar=NA)

# Convert warranty length from years to days
   WrntDays <- round(Wrnty * 365.25)

# PART DETAILS
# ============
# Work out rows from partFrq that need to be read.
   start <- min(index.rowcol(partFrq, partFrq==ClaimsFrom, "rows"))
   end <- max(index.rowcol(partFrq, partFrq==ClaimsTo, "rows"))

# Set index for output data frame to 1.
   i <- 1
   for( j in start : end )
   {  thisPart <- names(partFrq[j])    # This gives the part id

# If this entry in partFrq has no part id, go on to next value of i.
      if( thisPart=="" ) next

# Fill the output data frame with part, number and cost information.
      output[i,"Part"] <- thisPart
      output[i,"Claims"] <- partFrq[j]
      output[i,"CmpCst"] <- mean(CompDF$cost[CompDF$part==thisPart])
      output[i,"CmpCstSD"] <- stdev(CompDF$cost[CompDF$part==thisPart])

# Choose partname that occurs most frequently.
   partsTbl <- table(CompDF$partname[CompDF$part==thisPart])
   output[i,"Name"] <- names(partsTbl[partsTbl==max(partsTbl)])
# Issue a warning if part has no name.
   if( output[i,"Name"]=="" )
      warning("Inputted part has no name. May be more than one part.")

# SURVIVAL DATA FRAME AND EXPONENTIAL MODEL
# ========================================
# Create a survival dataframe and fit an exponential model.
      SurviveDF <- fSurvive(CompDF[CompDF$part==thisPart,], SalesDF)
      ExpMdl <- survReg(Surv(age2 + 1, status, type = "right") ~ 1, data = SurviveDF,
         na.action = na.exclude, dist = "exponential", control = list(maxiter = 30,
         rel.tolerance = 1e-005, failure = 1))

# Assign failure rate and variance from the exponential parameter (lambda = failure rate).
      output[i,"Rate"] <- exp(-ExpMdl$coeff)     # Daily failure rate
      RateVar <- ExpMdl$var*exp(-2*ExpMdl$coeff)
      output[i,"RateSD"] <- sqrt(RateVar)
      output[i,"NumRep"] <- output[i,"Rate"] * WrntDays * Production

# Fill the output data frame with warranty cost information.
      output[i,"WrntCst"] <- output[i,"CmpCst"] * output[i,"Rate"] * WrntDays

# For one-claim components, assign a value for variance of warranty cost based on variance of
# failure rate. Variance of cost does not exist when there is only one cost. Take cost as a
# constant.
      if(output[i,"Claims"]==1)
         output[i,"WrntVar"] <- (output[i,"CmpCst"]^2* RateVar) * WrntDays^2
```

```
      else
         output[i,"WrntVar"] <- ( output[i,"CmpCst"]^2 * RateVar +
            output[i,"Rate"]^2 * output[i,"CmpCstSD"]^2 +
            RateVar * output[i,"CmpCstSD"]^2 ) * WrntDays^2

      output[i,"WrntSD"] <- sqrt( output[i,"WrntVar"] )

# Increment i.
      i <- i + 1
   }

# Return the data frame.
   output
}
```

C.2 Exponential Fail Rate (*FailRate.ssc*)

```
#          FAILURE RATES FOR DIFFERENT FAILURE TIMES

#------------
# DESCRIPTION
#------------
# PURPOSE
# =======
# This function simulates component failure, with number of claims ranging from 1 to maxClaims
# (input parameter). Three secenarios are depicted:
# (1) all claims are made at the start of the warranty period;
# (2) all claims are made at the end of the warranty period;
# (3) claims are evenly distributed throughout the warranty period.

# INPUT
# =====
# The maximum number of claims for which the failure rate is required (default = 20).
# The length of the warranty in years (default = 3).
# The sales data frame of sold vehicles (default=sales97).

# OUTPUT
# ======
# A matrix containing the number of claims and the fail rates for each of the above three
# scenarios.

# -------
# PROGRAM
# -------
"fFailRate"<-
function(maxClaims=50, warranty =3, SalesDF=sales97)
{
# Convert the warranty to days.
   wrnty <- round(warranty * 365.25)

# Calculate how many vehicles produced.
   production <- length(SalesDF$VIN)
# Create matrix to hold failure rates.
   out <- matrix(data=NA, nrow=maxClaims, ncol=5,
          dimnames=list(1:maxClaims, c("NumClaims","StrRate", "EndRate","UnfmRate","NumRep") )
)

   for( i in 1:maxClaims )    # i is the number of failures
   {

# FAILURE AT START OF WARRANTY PERIOD
# ===================================
# Set age at failure.
      failAge <- 1

# Create dataframe of component failures.
      compDF <- data.frame(VIN=1:i, sale=rep(dates("1/1/97"),i),
```

```
        repr=rep(dates("1/1/97")+failAge,i), age2=rep(failAge,i),
        dist2=rep(NA,i), part=rep("a1",i), partname=rep("part a1",i), cost=rep(NA,i))
     SurviveDF <- fSurvive(compDF, SalesDF)
# Since the failed vehicle(s) have made-up VINs, fSurvive() could not find them in SalesDF, so
# an additional i records were written to SurviveDF which need to be removed. These can be
# removed from the end of the data frame.
     SurviveDF <- remove.row(target=SurviveDF, start.row=production+i+1, count=i)

# Fit exponential model to data.
     expMdl <- survReg(Surv(age2 + 1, status, type = "right") ~ 1, data = SurviveDF,
        na.action = na.exclude, dist = "exponential", control = list(maxiter = 30,
        rel.tolerance = 1e-005, failure = 1))

# Assign number of claims.
     out[i,"NumClaims"] <- i

# Assign failure rate when all claims at start of warranty.
     out[i,"StrRate"] <- 1/exp(expMdl$coeff)    # Daily failure rate

# FAILURE AT END OF WARRANTY PERIOD
# ================================
# Set age at failure to 1 less than warranty period.
     failAge <- wrnty-1
# Create dataframe of component failures.
     compDF <- data.frame(VIN=1:i, sale=rep(dates("1/1/97"),i),
        repr=rep(dates("1/1/97") + failAge,i), age2=rep(failAge,i),
        dist2=rep(NA,i), part=rep("a1",i), partname=rep("part a1",i), cost=rep(NA,i))
     SurviveDF <- fSurvive(compDF, SalesDF)
# Since the failed vehicle(s) have made-up VINs, fSurvive() could not find them in SalesDF, so
# an additional i records were written to SurviveDF which need to be removed. These can be
# removed from the end of the data frame.
     SurviveDF <- remove.row(target=SurviveDF, start.row=production+i+1, count=i)

# Fit exponential model to data.
     expMdl <- survReg(Surv(age2 + 1, status, type = "right") ~ 1, data = SurviveDF,
        na.action = na.exclude, dist = "exponential", control = list(maxiter = 30,
        rel.tolerance = 1e-005, failure = 1))

# Assign failure rate when all claims at end of warranty.
     out[i,"EndRate"] <- 1/exp(expMdl$coeff)    # Daily failure rate

# FAILURE UNIFORMLY DISTRIBUTED DURING WARRANTY PERIOD
# ==================================================
# Set uniformly distributed failure age as a vector of length i.
     failAge <- seq(from=round(wrnty/(i+1)), by=round(wrnty/(i+1)), length=i)

# Create dataframe of component failures.
     compDF <- data.frame(VIN=1:i, sale=rep(dates("1/1/97"),i),
        repr=dates("1/1/97")+failAge, age2=failAge,
        dist2=rep(NA,i), part=rep("a1",i), partname=rep("part a1",i), cost=rep(NA,i))
     SurviveDF <- fSurvive(compDF, SalesDF)
# Since the failed vehicle(s) have made-up VINs, fSurvive() could not find them in SalesDF, so
# an additional i records were written to SurviveDF which need to be removed. These can be
# removed from the end of the data frame.
     SurviveDF <- remove.row(target=SurviveDF, start.row=production+i+1, count=i)

# Fit exponential model to data.
     expMdl <- survReg(Surv(age2 + 1, status, type = "right") ~ 1, data = SurviveDF,
        na.action = na.exclude, dist = "exponential", control = list(maxiter = 30,
        rel.tolerance = 1e-005, failure = 1))

# Assign failure rate when claims uniformly distributed throughout the warranty.
     out[i,"UnfmRate"] <- 1/exp(expMdl$coeff)      # Daily failure rate

# ESTIMATED NUMBER OF REPLACEMENTS
# ================================
# Using the rate obtained from uniformly distributed failures, estimate the number of
# replacements during the warranty period
```

```
      out[i,"NumRep"] <- out[i,"UnfmRate"] * wrnty * production
   }
# Return matrix.
   out
}
```

C.3 Exponential Modelling with Grouping (*CostExpGp.ssc*)

```
# POINT & INTERVAL WARRANTY COST ESTIMATES USING A GROUPED EXPONENTIAL MODEL

#------------
# DESCRIPTION
#------------
# PURPOSE
# =======
# The exponential distribution is used to model failure of components with up to claimsTo
# during the warranty period. An estimate of the warranty cost and an upper limit are made.

# OUTLINE
# =======
# This program is in three parts:
# 1.   The number of components, the sum of the average costs of repair of these components and
#      an upper limit on these sums are obtained.
# 2.   The failure rate and an upper limit for it are obtained. These statistics depend on the
#      part chosen because there will be variations on the timing of failures.
# 3.   The expected warranty cost and an upper limit for it are obtained.

# INPUT
# =====
# 1. claimsTo = The max number of claims to be modelled by the exponential distribution.
#  Should be no more than 20 (the default).
# 2. confLvl = The two-sided confidence limit for the failure rate, cost of repair and warranty
#  cost. Default = 0.95.
# 3. warranty = The warranty length in years. Default = 3.
# 4. Plot = Logical variable indicating whether graphs of the costs are required. Default is F.

# OUTUT
# =====
# A data frame containing the number of claims (NumClaims), the number of components with 1 to
# claimsTo (NumComp), the sum of the average cost of repair (TotCst), together with its
variance
# (TotCstVar), the fail rate (Rate) and its standard deviation (RateSD), the expeceted number
of
# repairs predicted by the model (NumRep), the warranty cost Wrnt) and its variance (WrntVar)
# and a flag indicating whether the components were blank (BlnkPrt).

# -------
# PROGRAM
# -------
"fCstExpGp"<-
function(input=compAge, claimsFrom=1, claimsTo=20, confLvl=0.95, warranty=3, production=30138,
    Plot=F)
{

# CHECK AND INITIALISE VARIABLES
# ==============================
# Convert warranty length from years to days
   wrntDays <- round(warranty * 365.25)

# Check the max number of claims that has been passed in.
   if(claimsTo > 20)
      warning("Components have more than 20 claims. Exponential model may not be suitable")

# Check that claimsFrom < or = claimsTo
   if(claimsFrom > claimsTo)
      stop("claimsFrom is greater than claimsTo")
```

```
# -----------------------------------------------------
# 1. NUMBER OF COMPONENTS, SUM OF AVERAGE COST AND VARIANCE
# -----------------------------------------------------

# 1.1 CREATE A LIST OF part FREQUENCIES
# ===================================
# Creates a table of parts that are in the input data frame & sorts these if it doesn't already
# exist.
   if( !exists("partFrq") )
      partFrq <- sort( table(input$part), na.last=F )

# 1.2 CREATE A DATA FRAME FOR COMPONENTS WITH UP TO claimsTo
# ======================================================
# Create data frame to hold the data.
   output <- data.frame(NumClaims=NA,NumComp=NA,TotCst=NA,TotCstVar=NA,
                Rate=NA,RateSD=NA,NumRep=NA,Wrnt=NA,WrntVar=NA)

# Set up variables.
# i is the index of output.
   i <- 1

# j is the number of claims.
   if( is.element(claimsFrom, partFrq) )
      j <- claimsFrom
   else
      j <- min(partFrq[partFrq > claimsFrom])

   partFrqTbl <- table(partFrq)

# k is the index of partFrqTbl for the current number of claims (j).
   start <- index.rowcol(partFrqTbl, names(partFrqTbl)==j, which="rows")

# Fill the output data frame.
#  start <- min(index.rowcol(partFrq, partFrq==claimsFrom, "rows"))
   end <- max(index.rowcol(partFrqTbl, as.numeric(names(partFrqTbl))<=claimsTo, "rows") )

   for( k in start:end )
#  while(j <= claimsTo)
#  for(i in claimsFrom:claimsTo)
   {
# Parts is a temporary vector (changes with each i) to hold parts with frequency i.
      Parts <- names(partFrq[partFrq==j])

# If there are no parts with frequency = i, go on to next value of i.
#     if( is.null(Parts) )
#        next

# If Parts contains a component without a name, then it has to be dealt with separately.
      if( is.element("",Parts) & length(Parts==1) )
      {  #k <- k + 1
         j <- as.numeric(names(partFrqTbl[k+1]))
         next
      }
      if( is.element("",Parts) & length(Parts>1) )
      {  Parts <- sort(Parts)
         Parts <- Parts[2:length(Parts)]
      }
# costDF is a temporary data frame (changes with each i to hold parts with frequency i.
      costDF <-
         sort.col(target=input[is.element(input$part,Parts),c("part","partname","cost")],
         columns.to.sort="<ALL>", columns.to.sort.by=c("part","cost"))

# Fill the output data frame.
      output[i,"NumClaims"] <- j
      output[i,"NumComp"] <- length(unique(Parts))
      output[i,"TotCst"] <- sum(costDF$cost)/j
```

```
# Variance of the sum of variables is the sum of the individual variances.
      CstVar <- tapply(costDF$cost, costDF$part, var)
      output[i,"TotCstVar"] <- sum(CstVar)

# If a cost density plot is required, run this:
      if(Plot==T)
      {  plot(density(costDF$cost, from=0),xlab="cost", ylab="", type="l", Page="New")
         title(paste(Cost: Part,"costDF$cost") )
      }

# ------------------------------------------------------------------
# 2. FAILURE RATE AND STANDARD DEVIATION: UNIFORMLY DISTRIBUTED CLAIMS
# ------------------------------------------------------------------
# Set uniformly distributed failure age as a vector of length j (the number of claims).
      failAge <- round(seq(from=wrntDays/(j+1), by=wrntDays/(j+1), length=j))

# Create dataframe of component failures.
      compDF <- data.frame(VIN=1:j, sale=rep(dates("1/1/97"),j),
         repr=dates("1/1/97")+failAge, age2=failAge,
         dist2=rep(NA,j), part=rep("a1",j), partname=rep("part a1",j), cost=rep(NA,j))
      SurviveDF <- fSurvive(compDF, sales97)
      expMdl <- survReg(Surv(age2 + 1, status, type = "right") ~ 1, data = SurviveDF,
         na.action = na.exclude, dist = "exponential", control = list(maxiter = 30,
         rel.tolerance = 1e-005, failure = 1))

# Assign failure rate and variance from the exponential parameter (lambda = failure rate).
      output[i,"Rate"] <- exp(-expMdl$coeff)      # Daily failure rate
      RateVar <- expMdl$var*exp(-2*expMdl$coeff)
      output[i,"RateSD"] <- sqrt(RateVar)

# Obtain the number of repairs during the warranty.
      output[i,"NumRep"] <- output[i,"Rate"] * wrntDays * production

# ------------------------------------------
# 3. EXPECTED WARRANTY COST AND ITS VARIANCE
# ------------------------------------------
# Point estimate of warranty cost per vehicle.
      output[i,"Wrnt"] <- output[i,"Rate"]*output[i,"TotCst"]*wrntDays

# Confidence limits on warranty cost.
      if(j==1)
# Variance of warranty cost.
         output[i,"WrntVar"] <- sum(costDF$cost^2)* RateVar * wrntDays^2

      else
      {
# Need the sum of squares of individual repair costs.
      MeanCst <- tapply(costDF$cost, costDF$part, mean)

# Variance of warranty cost.
         output[i,"WrntVar"] <-
            ( sum(MeanCst^2)* RateVar + output[i,"Rate"]^2 * output[i,"TotCstVar"] +
            RateVar * output[i,"TotCstVar"] ) * wrntDays^2
      }
      i <- i + 1
#     k <- k + 1
      j <- as.numeric(names(partFrqTbl[k+1]))
   }

# Data frame may contain rows of NAs if there were no parts with a particualar number of claims
# up to claimsTo. Select the non NA rows.
   output <-select.rows(output, !is.na(output[,1]) )

# Return the data frame.
   output
}
```

C.4 Renewal Function Solver (*Renew.ssc*)

```
#              RENEWAL FUNCTION CALCULATOR FOR WEIBULL DISTRIBUTION

#------------
# DESCRIPTION
#------------
# OVERVIEW
# ========
# This function uses numerical methods as set out in Xie(1989) to return a matrix of expected
# number of renewals, M(i), using a Weibull distribution. Beta and eta are passed in as
# parameters.

# INPUT
# =====
# Values for the Weibull parameters eta and beta, the warranty length in years, and the step
# size, h, for the numerical calculations. The default value of h is 2 for accurate and speedy
# (calculations takes about 10 secs; h = 1 takes about 1 min). All four input parameters are
# single numeric values, not vectors.

# OUTPUT
# ======
# The expected number of renewals by the end of the warranty period.

# -------
# PROGRAM
# -------
"fRenew" <-
function(beta, eta, warranty=3, h=2)
{
# INITIALISE VARIABLES
# ====================
# Convert warranty length from years to days
   warranty <- round(warranty * 365.25)

# The values of these variables may be changed here. (Could set them up as function
parameters.)
# The function works for 5,000 intervals (t/h), but not 10,000 intervals. Eg, could have
# warranty = 1000 and h = 0.2.
# Note: The function fCompCst, which calls this function, didn't work when h was set to 1.
   numIntrvls <- floor(warranty/h) + 1

# SET VALUES OF F(t)
# ==================
# Set values of F(t) at interval ends.
   intrvls <- seq(from=0, to=warranty, by=h)
   Fend <- pweibull(intrvls, shape=beta, scale=eta)   # Cumulative Weibull function

# Set values of F(t) at interval midpoints.
   halfIntrvls <- seq(from=h/2, to=warranty, by=h)
   Fmid <- pweibull(halfIntrvls, shape=beta, scale=eta)
# Note: Fmid[1] = F(t=h/2).

# CALCULATE VALUES OF M(t)
# ========================
# Initialise vector M to 0. Thus M[1] = M(t=0) is set to 0.
    M <- vector(mode="numeric", length=numIntrvls)

# Set the value of M[2], which is M(t=h).
   M[2] <- Fend[2]/(1 - Fmid[1])

# Set the values of the rest of M[i].
   for (i in 3:numIntrvls)
   {
      sum <- 0
      for (j in 1:(i-2))
         sum <- sum + Fmid[i-j] * ( M[j+1] - M[j] )
      M[i] <- ( Fend[i] + sum - Fmid[1] * M[i-1] )/(1 - Fmid[1])
```

```
   }

# Return M(warranty):
   M[numIntrvls]
}
```

C.5 Weibull Modelling (*CostWbl.ssc*)

```
#  POINT & INTERVAL WARRANTY COST ESTIMATES USING A WEIBULL MODEL

#------------
# DESCRIPTION
#------------
# PURPOSE
# =======
# Point estimates and standard deviations of:
# 1. expected number of repairs of components, using a Weibull model, and
# 2. the cost of repairing a component.
# These are used to obtain an estimate of the warranty cost and its upper confidence limit.

# OUTLINE
# =======
# This function should be called by running the script runCostWbl. Alternatively, it can be
# called using the function tapply() or by(), so that analysis can be carried out by component.

# A survival data frame is created for a part and Weibull model fitted to the data using
# survReg(). Estimates of beta and eta are then extracted, and passed onto fRenew() to obtain
# estimates of the expected number of renewals during the warranty period.

# INPUT
# =====
# 1. CompDF = A subset of the data frame compAge, containing one part only. This can be done
# by calling this function using tapply() or by().
# 2. SalesDF = A data frame of sales. Default = sales97.
# 3. Wrnty = The warranty length in years. Default = 3.
# 4. confLvl = The confidence level for the parameter estimates. Default = 0.95.

# OUTPUT
# ======
# The output is a dataframe containing the following:
# Part details: part id (Part), part name (Name), number of claims (Claims);
# Repair cost details: cost of component replacement (CmpCst) and its SD (CmpCstSD);
# Weibull model details: parameter estimates (Beta, Eta), the expected number of renewals
#   (Renew) and its SD (RenewSD), the expected number of repairs during the warranty period
#   (NumRep), and the expected warranty cost (WrntCst) and its SD (WrntCstSD) and variance
#   (WrntCstVar).

# -------
# PROGRAM
# -------
"fCstWbl" <-
function(CompDF, SalesDF=sales97, Wrnty=3, Production=30138, confLevel=0.95)
{
# CHECK INPUT PARAMETERS
# ======================
   if(length(CompDF$VIN) == 0)
      stop(message="No components with that part number.")
   if(length(table(CompDF$part)) != 1)
      stop(message="This parameter 'IndCompDF' must contain only one part.")
   if(length(SalesDF$VIN) == 0)
      stop(message="The inputted sales database is empty.")
   if(Wrnty <= 0)
      stop(message="The parameter 'Wrnty' (warranty period in years) should be positive.")
   if(Production <= 0 )
      stop(message="The parameter 'Production' should be positive.")

# PART DETAILS
# ============
```

```
# Obtain part id from inputted data
   Part <- CompDF$part[1]

# Choose partname that occurs most frequently.
   partsTbl <- table(CompDF$partname)
   Name <- names(partsTbl[partsTbl==max(partsTbl)])
# Issue a warning if part has no name.
   if( Name=="" )
      warning("Inputted part has no name. May be more than one part.")

# Calculate the number of claims on that part.
   Claims <- count.rows(CompDF)

# Obtain mean cost of repairing a component.
   CmpCst <- mean(CompDF$cost, na.rm=T)

# Obtain the standard error of the cost of repair.
   CmpCstVar <- var(CompDF$cost, na.method="omit")
   CmpCstSD <- sqrt(CmpCstVar)

# SURVIVAL DATA FRAME AND WEIBULL MODEL
# ====================================
# Obtain a survival data frame for a component.
   SurviveDF <- fSurvive(CompDF, SalesDF)

# Weibull survival model.
   WblMdl <- survReg(Surv(age2 + 1, status, type = "right") ~ 1, data = SurviveDF,
      na.action = na.exclude, dist = "weibull", scale=0,
      control=list(maxiter = 50, rel.tolerance=1e-005, failure=1))

# POINT ESTIMATE OF THE NUMBER OF RENEWALS AND WARRANTY COST
# =========================================================
# Obtain the expected number of renewals (point estimate).
   Renew <- fRenew(beta=1/WblMdl$scale, eta=exp(WblMdl$coeff), warranty=Wrnty, h=1)

# Obtain the expected warranty cost.
   WrntCst <- Renew * CmpCst

# INTERVAL ESTIMATES OF THE EXPECTED NUMBER OF RENEWALS AND WARRANTY COST
# ======================================================================
# Obtain a confidence region for the Weibull parameters.
   degFree <- length(SurviveDF$VIN[SurviveDF$status==1])-3
   region <- fConfRgn(WblMdl, degFree, confLvl=confLevel, graph=F)

# If fConfRgn returns "singular", WblMdl generated a singular matrix, so no inverse exits.
# Therefore, confidence intervals cannot be obtained.
   if(region[[1]][[1]][1] == "singular")   # This extracts the first element
   {  RenewVar <- NA
      WrntCstVar <- NA
      WrntCstSD <- NA
   }
   else          # if inverse matrix exists
   {
# Obtain the expected number of renewals for these points.
      num <- length(region[[1]]$x)
      renRegn <- vector(mode="numeric", length=num)
      renRegn <- rep(NA,num)

# For loop needed here as fRenew() requires its parameters as single values, not vectors.
      for(i in seq(from=1, to=num, by=1) )
         renRegn[i] <- fRenew(beta=region[[1]]$x[i], eta=region[[1]]$y[i], warranty=3, h=2)

      RenewVar <- sum( (renRegn-Renew)^2/length(renRegn[!is.na(renRegn)]), na.rm=T )

# Variance of warranty cost.
      WrntCstVar <- Renew^2*CmpCstVar + CmpCst^2*RenewVar + CmpCstVar * RenewVar
      WrntCstSD <- sqrt(WrntCstVar)
   }
```

```
# OUTPUT
# ======
# Create output.
   output <- data.frame(
       Part = Part,
       Name = Name,
       Claims = Claims,
       CmpCst = CmpCst,
       CmpCstSD = sqrt(CmpCstVar),
       Beta = 1/WblMdl$scale,
       Eta = exp(WblMdl$coeff),
       Renew = Renew,
       RenewSD = sqrt(RenewVar),
       NumRep = Renew * Production,
       WrntCst = WrntCst,
       WrntCstSD = WrntCstSD,
       WrntCstVar = WrntCstVar,
   )
#  output <- convert.col.type(target=output, column.spec=list("part","name"),
#      column.type="character")

# Return output as a data frame.
   output
}
```

C.6 Confidence Region (*ConfRgn.ssc*)

```
#      CONFIDENCE REGION (ELLIPSOID)

#------------
# DESCRIPTION
#------------
# OVERVIEW
# ========
# Letting 'WblMdl' be a Weibull model of survival data, point estimates of the Weibull
# parameters eta (scale) and beta (shape) are obtained:
#      beta = 1/WblMdl$scale.
#      eta = exp(WblMdl$coeff).
# This function returns a confidence region for the Weibull parameters using the variances of
# WblMdl$coeff and log(WblMdl$scale) from the inputted survival model 'WblMdl'

# DETAILS
# =======
# Reference: Cook and Weisberg (1994) pp 218-9.
# For a two-predictor model (T written for theta):
#   y|x = a0 + a1.x1 + a2.x2 + error
# with independendent normal errors having mean 0 and constant variances, a joint (1-alpha)100%
# confidence region for T = transpose(a1, a2) is the set of all values of the 2x1 vector T that
# satisfies the inequality:
#   transpose(T-T_hat).inverse(var(T_hat)).(T-T_hat) < or = 2.F(1-alpha ,2,n-3)
# where F(1-alpha ,2,n-3) is the percentage point of the F-distribution with 2 and n-3 degrees
# of freedom that leaves an area of alpha under the right tail. The points that satisfy this
# inequality fall inside an ellipse with centre at T_hat. In the Weibull parameterisation, we
# have a1 = eta and a2 = beta.

# A grid of 100 values (xGrid) between beta +/- 3 standard deviations is generated and
# a grid of 100 values (yGrid) between eta +/- 3 standard deviations is also generated.
# Matrix multiplication is performed on each point (xGrid, yGrid) in turn. Thus, a 100x100
# matrix is generated. Each element of the matrix is compared to 2.F(1-alpha ,2,n-3) to find
the
# points that are within the confLvl confidence region.

# INPUT
# =====
# WblMdl is a survReg object. It can be a Weibull fit to the survival data frame.
# numFail is length(SurviveDF$VIN[SurviveDF$status==1]). It is needed to calculate the
# degrees of freedom 2 parameter for the F-distribution. The formula is
# df2 = length(SurviveDF$VIN[SurviveDF$status==1]) - 3. The -3 is done in this function.
```

```
# graph is a logical parameter indicating whether a graph should be produced (default=F).

# -------
# PROGRAM
# -------
"fConfRgn" <- function(WblMdl, numFail, confLvl=0.95, graph=F)
{
# Obtain eta and beta estimates from WblMdl$coeff and WblMdl$scale, and estimates of their
# variances and the variance/covariance matrix.
   beta <- 1/WblMdl$scale
   varBeta <- beta^2 * WblMdl$var[2,2]

   eta <- exp(WblMdl$coeff)
   varEta <- eta^2 * WblMdl$var[1,1]

   covBetaEta <- -exp(WblMdl$coeff)*WblMdl$var[1,2]/WblMdl$scale

# The variance/covariance matrix is:
   varCov <- matrix(c(varBeta, covBetaEta, covBetaEta, varEta), nrow=2,ncol=2)

# Obtain the inverse of varCov manually. Solve doesn't seem to work.
# If varCov is singular, return a warning message and exit this function returning -999.
   if(varCov[1,1]*varCov[2,2]-varCov[1,2]*varCov[2,1] == 0)
   {  warning(message="Singular matrix; inverse does not exist.")
      return("singular")
   }

# If varCov is not singular, evaluate its inverse:
# Generate grid points for the contour plot. Exclude any negative values of beta or eta by
# limiting the start of the grid to 1e-5, as 0 is not a valid value for either parameter.
      betaGrid <- seq(max(beta - 3*sqrt(varBeta),1e-5), beta + 3*sqrt(varBeta), length = 100)
      etaGrid <- seq(max(eta - 3*sqrt(varEta),1e-5), eta + 3*sqrt(varEta), length=100)

# Generate a 100x100 matrix of zeros.
      grid <- matrix(data=0, nrow=100, ncol=100)

      inverse <-
         matrix(c(varCov[2,2],-varCov[1,2],-varCov[2,1],varCov[1,1]),nrow=2,ncol=2,byrow=T)/
         (varCov[1,1]*varCov[2,2]-varCov[1,2]*varCov[2,1])

#   else inverse <- solve(varCov)
# Solve seems to to think some matrices are singular when the discriminant != 0.

# Fill the matrix by performing matrix multiplication on every point (xGrid, yGrid).
   for (j in 1:100)
   {  for (i in 1:100)
      {  theta <-  matrix(c(betaGrid[i] - beta, etaGrid[j] - eta), nrow=2, ncol=1)
         grid[i,j] <- t(theta) %*% inverse %*% theta
      }
   }

# Generate a contour plot of points showing the inputted (confLvl) confidence region using the
# F-distribution. Set the contour levels.
   FLevel <- 2*qf(confLvl, 2, numFail-3)

# If pararmeter 'graph' = T, create a contour plot.
   if(graph==T)
   {  graphsheet()
      contour(betaGrid, etaGrid, grid, levels = FLevel, xlab="beta", ylab="eta")
      title("(beta, eta) Confidence Region")
      points(1/WblMdl$scale, exp(WblMdl$coeff), type="p")
   }
# The following returns coordinates on (betagrid, etagrid) where the contour level (grid) =
# the confidence level for the F distibution (Flevel) set above. These points can be passsed to
# fRenew() to generate values of M(t), to obtain bounds on M(t).

contour(betaGrid, etaGrid, grid, levels = FLevel, save=T, plotit=F)
}
```

C.7 Combined Exponential and Weibull Modelling Function (*CostExpWbl.ssc*)

```
#       POINT & INTERVAL WARRANTY COST ESTIMATES USING EXPONENTIAL AND WEIBULL MODELS

#------------
# DESCRIPTION
#------------
# PURPOSE
# =======
# This function obtains point estimates and standard deviation of the following:
# 1. The mean cost of repairing a component;
# 2. The failure rate using an exponential model;
# 3. The values of the parameters of a Weibull model;
# 4. The expected number of renewals of components using a Weibull model;
# 5. The expected number of repairs during the warranty period for a year's production for each
#  of the exponential and Weibull models;
# 6. The expected warranty cost of replacing a component for each model.

# OUTLINE
# =======
# This function should be called using the script callCostExpWbl.

# For each part passed into this function, the part details are extracted. Then a survival data
# frame is created for each part and both exponential and Weibull models are fitted to the data
# using survReg(). Using the models, the quantities listed in Purpose are estimated.

# INPUT
# =====
# 1.   CompDF = A subset of the data frame compAge, containing one part only. This can be done
#      by calling this function using  the script callCostExpWbl.
# 2.   SalesDF = A data frame of sales. Default = sales97.
# 3.   Wrnty = The warranty length in years. Default = 3.
# 4.   ConfLvl = The confidence level for the parameter estimates. Default = 0.95.
# 5.   Production = The number of vehicles produced in a year. Default = 30138.
# 6.   MinClmsWbl = The min number of claims to be modell by the Weibull distribution.
#      Default = 20.

# OUTPUT
# ======
# The output is a dataframe containing the following:
# Part details: part id (Part), part name (Name), number of claims (Claims);
# Repair cost details: cost of component replacement (CmpCst) and its SD (CmpCstSD);
# Expoential model details: the failure rate (Rate) and its SD (RateSD), the expected number of
#  repairs during the warranty period (NumRepExp), and the expected warranty cost (WrntExp)
#  and its SD (WrntExpSD) and variance (WrntExpVar);
# Weibull model details: parameter estimates (Beta, Eta), the expected number of renewals
#  (Renew) and its SD (RenewSD), the expected number of repairs during the warranty period
#  (NumRepWbl), and the expected warranty cost (WrntWbl) and its SD (WrntWblSD) and
#  variance (WrntWblVar).
# Model comparison: Loglikelihood of the exponential fit (LoglikExp) and the Weibull fit
#  (LoglikWbl) and an indication as to which is the maximum of these (BestFit).

# -------
# PROGRAM
# -------
"fCstExWb" <-
function(CompDF, SalesDF=sales97, Wrnty=3, Production=30138, ConfLevel=0.95, MinClmsWbl=20)
{
# CHECK INPUT PARAMETERS
# ======================
   if(length(CompDF$VIN) == 0)
      stop(message="No components with that part number.")
   if(length(CompDF$VIN) < 20)
      warning("Less than 20 claims on that component. Weibull model may be unsuitable.")
   if(length(table(CompDF$part)) != 1)
      stop(message="This function requires the input data frame to have only one part.")
   if(length(SalesDF$VIN) == 0)
      stop(message="The inputted sales database is empty.")
```

```
   if(Wrnty <= 0)
      stop(message="The parameter 'Wrnty' (warranty period in years) should be positive.")
   if(Production <= 0 )
      stop(message="The parameter 'Production' should be positive.")

# Convert warranty to days.
   WrntDays <- round(Wrnty * 365.25)

# INITIALISE WEIBULL MODEL VARIABLES
# ==================================
# These variables will need to be NA if Weibull model is not fitted.
   Beta <- NA
   Eta <- NA
   Renew <- NA
   RenewSD <- NA
   NumRepWbl <- NA
   PtsTrim <- NA
   WrntWbl <- NA
   WrntWblSD <- NA
   WrntWblVar <- NA
   LoglikWbl <- NA
   BestFit <- NA
   SmlstSD <- NA
   WbSDLsCst <- NA

# PART DETAILS
# ============
# Obtain part id from inputted data
   Part <- CompDF$part[1]

# Choose partname that occurs most frequently.
   partsTbl <- table(CompDF$partname)
   Name <- names(partsTbl[partsTbl==max(partsTbl)])
# Issue a warning if part has no name.
   if( Name=="" )
      warning("Inputted part has no name. May be more than one part.")

# Calculate the number of claims on that part.
   Claims <- count.rows(CompDF)

# Obtain mean cost of repairing a component.
   CmpCst <- mean(CompDF$cost, na.rm=T)

# Obtain the standard error of the cost of repair.
   CmpCstVar <- var(CompDF$cost, na.method="omit")
   CmpCstSD <- sqrt(CmpCstVar)

# SURVIVAL DATA FRAME AND EXPONENTIAL MODEL
# =========================================
   SurviveDF <- fSurvive(CompDF, SalesDF)
   expMdl <- survReg(Surv(age2 + 1, status, type = "right") ~ 1, data = SurviveDF,
      na.action = na.exclude, dist = "exponential", control = list(maxiter = 30,
      rel.tolerance = 1e-005, failure = 1))
   LoglikExp <- expMdl$loglik[1]

# Assign failure rate and variance from the exponential parameter (lambda = failure rate).
   Rate <- as.numeric( exp(-expMdl$coeff) )      # Daily failure rate
   RateVar <- as.numeric( expMdl$var*exp(-2*expMdl$coeff) )
   RateSD <- sqrt(RateVar)
   NumRepExp <- Rate * WrntDays * Production

# Calculate warranty cost information.
   WrntExp <- CmpCst * Rate * WrntDays

# For one-claim components, assign a value for variance of warranty cost based on variance of
# failure rate. Variance of cost does not exist when there is only one cost. Take cost as a
# constant.
```

```
   if(Claims==1)
      WrntExpVar <- CmpCst^2 * RateVar * WrntDays^2
   else
      WrntExpVar <- (CmpCst^2 * RateVar + Rate^2 * CmpCstVar + RateVar*CmpCstVar) * WrntDays^2

   WrntExpSD <- sqrt(WrntExpVar)

# WEIBULL MODEL
# ============
# Do Weibull modelling only if number of claims is at least MinClmsWbl (inputted).
   if(Claims >= MinClmsWbl)
   {  wblMdl <- survReg(Surv(age2 + 1, status, type = "right") ~ 1, data = SurviveDF,
         na.action = na.exclude, dist = "weibull", scale=0,
         control=list(maxiter = 50, rel.tolerance=1e-005, failure=1))
      LoglikWbl <- wblMdl$loglik[1]

# ADJUST THE WEIBULL MODEL IF IT IS NOT A GOOD FIT
# ===============================================
# First arrange age2 in descending order to identify the highest values of age2.
      if(LoglikWbl < 1.5 *LoglikExp)
         ageDesc <- sort.col(target=CompDF$age2, columns.to.sort = T,
            columns.to.sort.by = T, ascending=F)

# If the Weibull fit is much poorer than the exponential fit, knock off records with the
biggest
# age2 until the Weibull loglikelihood is within 1.5 times that of the exponential likelihood.
      j <- 1    # Initialise the counter j
      while(LoglikWbl < 1.5 *LoglikExp & j < length(CompDF$VIN) )
      {  SurviveDF <- fSurvive(CompDF[CompDF$age2<ageDesc[j],], SalesDF)
         wblMdl <- survReg(Surv(age2 + 1, status, type = "right") ~ 1, data = SurviveDF,
            na.action = na.exclude, dist = "weibull", scale=0,
            control=list(maxiter = 30, rel.tolerance=1e-005, failure=1))
         LoglikWbl <- wblMdl$loglik[1]
         j <- j + 1
      }

# Assign the number of points trimmed in fitting the Weibull model.
      PtsTrim <- j - 1

# WEIBULL MODEL: POINT ESTIMATES
# ============================
# Weibull parameters.
      Beta <- 1/wblMdl$scale
      Eta <- exp(wblMdl$coeff)

# Obtain the expected number of renewals (point estimate).
      Renew <- fRenew(beta=Beta, eta=Eta, warranty=Wrnty, h=1)

# Obtain the expected warranty cost.
      WrntWbl <- Renew * CmpCst

# Expected number of repairs
      NumRepWbl <- Renew * Production

# WEIBULL MODEL: INTERVAL ESTIMATE - EXPECTED NUMBER OF RENEWALS AND WARRANTY COST
# ===============================================================================
# Obtain a confidence region for the Weibull parameters.
      degFree <- length(SurviveDF$VIN[SurviveDF$status==1])-3

# Only if degrees of freedom > 0 can fConfRgn() be called, as it uses F-distribution.
      if(degFree > 0)
      {  region <- fConfRgn(wblMdl, degFree, confLvl=ConfLevel, graph=F)

# If fConfRgn() returns "singular", wblMdl generated a singular matrix, so no inverse exits.
# Therefore, confidence intervals cannot be obtained. Should not get a singular matrix since
# the Weibull fit is adjusted above.
         if(region[[1]][[1]][1] != "singular")   # This extracts the first element
         {
```

```
# Obtain the expected number of renewals for these points.
            num <- length(region[[1]]$x)
            renRegn <- vector(mode="numeric", length=num)
            renRegn <- rep(NA,num)

# For loop needed here as fRenew() requires its parameters as single values, not vectors.
            for( i in 1:num )
               renRegn[i] <-
               fRenew(beta=region[[1]]$x[i], eta=region[[1]]$y[i], warranty=Wrnty, h=2)

            RenewVar <- sum( (renRegn-Renew)^2/length(renRegn[!is.na(renRegn)]), na.rm=T )
            RenewSD <- sqrt(RenewVar)

            WrntWblVar <- Renew^2*CmpCstSD^2 + CmpCst^2*RenewVar + CmpCstVar * RenewVar
            WrntWblSD <- sqrt(WrntWblVar)
         }
      BestFit <- if(LoglikWbl > LoglikExp) "Weibull" else "Exponential"
      SmlstSD <- if(WrntExpSD==WrntWblSD) "Equal"
               else
               {  if(WrntExpSD < WrntWblSD) "Exponential"
                  else "Weibull"
               }
      WbSDLsCst <- if(WrntWblSD < WrntWbl) "Yes" else "No"
      ExSDLsCst <- if(WrntExpSD < WrntExp) "Yes" else "No"
      }
   }

# OUTPUT
# ======
# Create output.
   output <- data.frame(
      Part=Part, Name=Name, Claims=Claims, CmpCst=CmpCst, CmpCstSD=CmpCstSD,
      Rate=Rate, RateSD=RateSD, NumRepExp=NumRepExp,
      WrntExp=WrntExp, WrntExpSD=WrntExpSD, WrntExpVar=WrntExpVar,
      Beta=Beta, Eta=Eta, Renew=Renew, RenewSD=RenewSD, NumRepWbl=NumRepWbl, PtsTrim=PtsTrim,
      WrntWbl=WrntWbl, WrntWblSD=WrntWblSD, WrntWblVar=WrntWblVar,
      LoglikExp=LoglikExp, LoglikWbl=LoglikWbl,
      BestFit=BestFit, SmlstSD=SmlstSD, WbSDLsCst=WbSDLsCst,
      ExSDLsCst=if(WrntExpSD < WrntExp) "Yes" else "No" )

# Return output as a data frame.
   output
}
```

C.8 Estimate of Total Warranty Cost (*WrntCstTot.ssc*)

```
#      TOTAL WARRANTY COST AND CONFIDENCE INTERVAL

#------------
# DESCRIPTION
#------------
# PURPOSE
# =======
# The purpose of this function is to obtain the total warranty cost, its variance and a
# confidence interval on the cost for a specified warranty period using (i) the exponential
# model, (ii) the exponential model with components grouped by claims frequency and (iii) the
# Weibull model and (iv) a restricted Weibull model, where if the warranty cost for 5 years
# under the Weibull model is a certain (inputted) factor times the warranty cost for 3 years,
# then the exponential model is used.

# OUTLINE
# =======
# The warranty cost and variance are worked out for each of the three models in turn. The
# calculation is done specifically for the data frames "costExpWbl" and "extWrnt", which have
# the required figures to be totalled.

# INPUT
# =====
```

```
# Input = a data frame of component warranty cost and variance. The two data frames to be used
# are "costExpWbl" (default) for a 3-year warranty and "extWrnt" for a 5-year warranty.
# Wrnty = the warranty length in years. Default = 3 years.
# ConfLvl = Confidence interval for the warranty cost. Default = 0.95.
# IncFctr = The factor at which the Weibull will be rejected if the 5-year warranty cost is
# that many times bigger than the 3-year warranty. Default = 5.

# OUTPUT
# ======
# A matrix with columns: Warranty Cost, Standard Deviation, Lower and Upper Confidence Limits;
# and rows: Exponential Model, Grouped Exponential Model, Weibull Modeland restricte Weibull
# model (if the warranty is 5 years).

# -------
# PROGRAM
# -------
"fWrntCst" <-
function(Input=costExpWbl, Wrnty=3, ConfLvl=0.95, IncFctr=5)
{
# INITIALISE VARIABLES
# ====================
   wrntDays <- round(Wrnty * 365.25)

# EXPONENTIAL MODEL
# =================
# Total warranty cost under the exponential model.
   TotCstExp <- sum(Input$WrntExp[Input$Part!=""], na.rm=T)
   CstBlankExp <-
      costExpWbl$CmpCst[costExpWbl$Part==""] * costExpWbl$Rate[costExpWbl$Part==""] * wrntDays
   TotCstExp <- TotCstExp + CstBlankExp

# Total warranty variance under the exponential model.
   TotCstExpVar <- sum(Input$WrntExpVar[Input$Part!=""], na.rm=T)
   CstBlankExpVar <-
      sum(compAge$cost[compAge$part==""]^2) * cstExpGp$RateSD[cstExpGp$NumClaims==1]^2 *
      wrntDays^2
   TotCstExpSD <- sqrt(TotCstExpVar + CstBlankExpVar)

# Lower confidence limit on warranty cost under the exponential model using normal
distribution.
   LwrWntExp <- max(qnorm( 0.5-ConfLvl/2, mean=TotCstExp, sd=TotCstExpSD), 0)

# Upper confidence limit on warranty cost of the exponential model using normal distribution.
   UprWntExp <- qnorm( 0.5+ConfLvl/2, mean=TotCstExp, sd=TotCstExpSD)

# GROUPED EXPONENTIAL MODEL
# =========================
# Total warranty cost under the exponential model with components grouped by claim frequency.
# In this calculation, "TotCstGpExp" is used, as it is the only place where the grouped
# exponential model is used. Since the data frame contains data for a 3-year warranty, it needs
# to be adjusted for the 5-year warranty.
# Note that there is not an entry in cstExpGp when part = "".
   TotCstGpExp <- sum(cstExpGp$Wrnt) * wrntDays / 1096
   TotCstGpExp <- TotCstGpExp + CstBlankExp

# Total warranty variance under the grouped exponential model.
   TotCstGpExpVar <- sum(cstExpGp$WrntVar, na.rm=T) * (wrntDays / 1096)^2
   TotCstGpExpSD <- sqrt(TotCstGpExpVar + CstBlankExpVar)

# Lower confidence limit on warranty cost under the grouped exponential model using normal
# distribution.
   LwrWntGpExp <- max(qnorm( 0.5-ConfLvl/2, mean=TotCstGpExp, sd=TotCstGpExpSD), 0)

# Upper confidence limit on warranty cost the grouped exponential model using normal
# distribution.
   UprWntGpExp <-  qnorm( 0.5+ConfLvl/2, mean=TotCstGpExp, sd=TotCstGpExpSD)
```

```
# WEIBULL MODEL
# =============
# Use exponential model for total warranty cost for parts with < 20 claims.
   TotCstExp20 <- sum(Input$WrntExp[Input$Claims<20 & Input$Part!=""], na.rm=T)

# Total warranty cost under the Weibull model for parts with at least 20 claims.
   TotCstWbl <- sum(Input$WrntWbl[Input$Claims>=20 & Input$Part!=""], na.rm=T)

# Total warranty cost with exponential model for parts with less than 20 claims, and Weibull
# model for parts with 20 or more claims.
   TotCstWbl <- TotCstWbl + TotCstExp20 + CstBlankExp

# Total warranty variance under the Weibull model for parts with at least 20 claims.
   TotCstWblVar <-
      sum(Input$WrntWblVar[Input$Claims>=20 & Input$Part!=""], na.rm=T)
   TotCstWblNAVar <-
      sum(Input$WrntExpVar[Input$Claims>=20 & Input$Part!="" & is.na(Input$WrntWblVar)],
      na.rm=T)
   TotCstExpVar20 <-
      sum(Input$WrntExpVar[Input$Claims<20 & Input$Part!=""], na.rm=T)
   TotCstWblSD <- sqrt(TotCstWblVar + TotCstWblNAVar + TotCstExpVar20 + CstBlankExpVar)

# Lower confidence limit on warranty cost under the Weibull model using normal distribution.
   LwrWntWbl <- max(qnorm(0.5-ConfLvl/2, mean=TotCstWbl, sd=TotCstWblSD), 0)

# Upper confidence limit on warranty cost the Weibull model using normal distribution.
   UprWntWbl <- qnorm( 0.5+ConfLvl/2, mean=TotCstWbl, sd=TotCstWblSD)

# OUTPUT FOR 3-YEAR WARRANTY
# ==========================
# Return this output of exponential and Weibull warranty costs, variances and confidence
# intervals if the warranty is 3 years.
   output <- matrix(
      data=c(TotCstExp,TotCstExpSD,LwrWntExp,UprWntExp,TotCstGpExp,TotCstGpExpSD,LwrWntGpExp,
         UprWntGpExp,TotCstWbl,TotCstWblSD,LwrWntWbl,UprWntWbl),
      nrow=3, ncol=4, byrow=T,
      dimnames=list( c("Exp","GpExp","Wbl"),c("WrntCst","WrntSD","LwrLmt","UprLmt") ) )

# ADJUSTED WEIBULL MODEL
# ======================
# Some parts are not modelled well by the Weibull model; their costs become too large at 5
# years, so instead, the exponential model will be used for those parts.
   if(Wrnty == 5)
   {  TotCstWblAdj <- sum(Input$WrntWbl[Input$Claims >= 20 & Input$Part != "" &
         Input$WrntWbl < IncFctr * costExpWbl$WrntWbl], na.rm=T)
      TotCstRjctWbl <- sum(Input$WrntExp[Input$WrntWbl >= IncFctr*costExpWbl$WrntWbl],na.rm=T)
      TotCstWblAdj <- TotCstExp20 + CstBlankExp + TotCstWblAdj + TotCstRjctWbl

# Total warranty variance under the adjusted Weibull model.
      TotCstWblAdjVar <-
         sum(Input$WrntWblVar[Input$Claims>=20 & Input$Part!="" &
         Input$WrntWbl<IncFctr*costExpWbl$WrntWbl], na.rm=T )
      TotCstWblNAVar <-
         sum(Input$WrntExpVar[Input$Claims>=20 & Input$Part!="" & is.na(Input$WrntWblVar)],
         na.rm=T)
      TotCstExpVar20 <-
         sum(Input$WrntExpVar[Input$Claims<20 & Input$Part!=""], na.rm=T)
      TotCstExpVarAdj <-
         sum(Input$WrntExpVar[Input$WrntWbl>=IncFctr*costExpWbl$WrntWbl], na.rm=T)
      TotCstWblAdjSD <- sqrt(
         TotCstWblAdjVar + TotCstWblNAVar + TotCstExpVar20 + TotCstExpVarAdj + CstBlankExpVar)

# Lower confidence limit on warranty cost under the Weibull model using normal distribution.
      LwrWntWblAdj <- max(qnorm(0.5-ConfLvl/2, mean=TotCstWbl, sd=TotCstWblSD), 0)

# Upper confidence limit on warranty cost the Weibull model using normal distribution.
      UprWntWblAdj <- qnorm( 0.5+ConfLvl/2, mean=TotCstWbl, sd=TotCstWblSD)
```

```
# OUTPUT FOR 5-YEAR WARRANTY
# ========================
# Return this output of exponential and Weibull warranty costs, variances and confidence
# intervals.
      output <- matrix(data=c(TotCstExp,TotCstExpSD,LwrWntExp,UprWntExp,
      TotCstGpExp,TotCstGpExpSD,LwrWntGpExp,UprWntGpExp,
         TotCstWbl,TotCstWblSD,LwrWntWbl,UprWntWbl,
         TotCstWblAdj,TotCstWblAdjSD,LwrWntWblAdj,UprWntWblAdj),
         nrow=4, ncol=4, byrow=T,
         dimnames=list(
            c("Exp","GpExp","Wbl","WblAdj"),c("WrntCst","WrntSD","LwrLmt","UprLmt")))
   }

# Return output
   output
}
```

C.9 Extended Warranty Cost (*ExtdWrnt.ssc*)

```
#      EXTENDED WARRANTY: POINT & INTERVAL WARRANTY COST ESTIMATES

#------------
# DESCRIPTION
#------------
# PURPOSE
# =======
# The aim of this function is to obtain point estimates and standard deviation of the following
# for each component for an extended warranty:
# 1. The expected number of repairs, warranty cost and variance using an exponential model.
# 2. The expected number of repairs, warranty cost and variance using a Weibull model.

# OUTLINE
# =======
# This function should be called using the script CallExtdWrnt.
# The function uses the model parameter values from the inputted data frame to work out the
# cost of an extended warrranty. Since the exponential model has a linear rate, the extended
# warranty cost is a linear extrapolation of the current warranty. For the Weibull model, a
# model has to be fitted. The log likelihood of the Weibull model is compared to the log
# likelihood of the exponential model. If the Weibull's log likelihood is not close to that of
# the expoential model, the model is adjusted by removing the oldest points, one at a time,
# until a good fit is achieved. Once a good fit is achieved, the Weibull model parameters are
# passed onto fRenew, and they, together with the variances, are passed onto fConfRgn, to
# obtain point and interval estimates of the warranty cost.
# If the 'WblMdl' parameter of fConfRgn could be broken up into its components, and if they
# could be included in ParmsDF, then they could be read by this function instead of fitting
# the model again, which would save a lot of processing time.

# INPUT
# =====
# 1.   ParmsDF = a dataframe containing the parameters of the exponential and Weibull models.
#      Default = costExpWbl.
# 2.   CompDF = a data frame of components' ages. Default = compAge.
# 3.   SalesDF = A data frame of sales. Default = sales97.
# 4.   StartRow = the first row of ParmsDF to be read. It takes too long to process all rows
#      of ParmsDF at once. Default = 1.
# 5.   EndRow = the last row of ParmsDF to be read. Default = 2693, which is the last row of
#      ParmsDF containing components with < 20 claims, after which, the Weibull model parameter
#      exist in ParmsDF, and are used in this function. Processing is fast without Weibull
#      modelling.
# 6.   Wrnty = the warranty length in years. Default = 5.
# 7.   Production = the number of vehicles produced in a year. Default = 30138.

# OUTPUT
# ======
# The output is a dataframe containing the expected number of repairs, warranty cost and
# variances for both the exponential and Weibull models.
```

```
#---------
# PROGRAM
#---------
"fExtWrnt" <-
function(ParmsDF=costExpWbl, CompDF=compAge, SalesDF=sales97, StartRow=1, EndRow=2693,
   Wrnty=5, Production=30138)
{
# CHECK INPUT PARAMETERS
# ======================
   if(Wrnty <= 0)
      stop(message = "The parameter 'Wrnty' (warranty period in years) should be positive.")
   if(StartRow > EndRow)
      stop("The parameter StartRow must be smaller than EndRow.")
   if(StartRow != round(StartRow, 0) & EndRow != round(EndRow, 0))
      stop("The parameters StartRow and EndRow must be positive whole numbers.")

# Convert warranty to days.
   WrntDays <- round(Wrnty * 365.25)

# SET UP OUTPUT AND INITIALISE WEIBULL MODEL VARIABLES
# ====================================================
# Select the required rows from ParmsDF, attach this data frame and set up output data frame.
   data <- ParmsDF[StartRow:EndRow, ]
   attach(data)
   output <- data[, c("Part", "Name", "Claims", "NumRepExp", "WrntExp", "WrntExpSD",
      "WrntExpVar", "NumRepWbl", "WrntWbl", "WrntWblSD", "WrntWblVar")]
   numRow <- length(StartRow:EndRow) # Clear all calculated variables to NA.
   output[,4:11] <- as.numeric( rep(NA, times = numRow) )

# EXPONENTIAL MODEL
# =================
# Calculate warranty cost for an extended warranty.
   output$WrntExp[1:numRow] <- CmpCst[1:numRow] * Rate[1:numRow] * WrntDays

# Calculate the expected number of repairs in 5 years predicted by the exponential model.
   output$NumRepExp[1:numRow] <- Rate[1:numRow] * WrntDays * Production

# For one-claim components, assign a value for variance of warranty cost based on variance of
# failure rate. Variance of cost does not exist when there is only one cost. Take cost as a
# constant.
   clms1 <- Claims == 1
   output$WrntExpVar[clms1] <- CmpCst[clms1]^2 * RateSD[clms1]^2 * 1826^2

# Records with blank names will be included below, but will be dealt with in wrntCst()
   output$WrntExpVar[!clms1] <- (CmpCst[!clms1]^2 * RateSD[!clms1]^2 + Rate[!clms1]^2 *
      CmpCstSD[!clms1]^2 + RateSD[!clms1]^2 * CmpCstSD[!clms1]^2) * WrntDays^2
   output$WrntExpSD <- sqrt(output$WrntExpVar)

# WEIBULL MODEL
# =============
   for(i in 1:numRow) {
# If Weibull model exists for current record, proceed.
      if(WrntWbl[i] != "NA") {
         DF <- CompDF[CompDF$part == Part[i], ]
         SurviveDF <- fSurvive(DF, SalesDF)
         wblMdl <- survReg(Surv(age2 + 1, status, type = "right") ~ 1, data = SurviveDF,
            na.action = na.exclude, dist = "weibull", scale = 0,
            control = list(maxiter = 50, rel.tolerance = 1e-005, failure = 1))
         lglikeWb <- wblMdl$loglik[1]

# ADJUST THE WEIBULL MODEL IF IT IS NOT A GOOD FIT
# ================================================
# First arrange age2 in descending order to identify the highest values of age2.
         if(lglikeWb < 1.5 * LoglikExp[i])
         {  ageDesc <- sort.col(target = DF$age2,
               columns.to.sort = T, columns.to.sort.by = T, ascending = F)
         }
```

```
# If the Weibull fit is much poorer than the exponential, knock off records with the biggest
# age2 until the Weibull loglikelihood is within 1.5 times that of the exponential likelihood.
        j <- 1 # Initialise the counter j
        while(lglikeWb < 1.5 * LoglikExp[i] & j < length(DF$VIN))
        {
            SurviveDF <- fSurvive(DF[DF$age2 < ageDesc[j], ], SalesDF)
            wblMdl <- survReg(Surv(age2 + 1, status, type = "right") ~ 1, data = SurviveDF,
                na.action = na.exclude, dist = "weibull", scale = 0,
                control = list(maxiter = 30, rel.tolerance = 1e-005, failure = 1))
            lglikeWb <- wblMdl$loglik[1]
            j <- j + 1
        }


# WEIBULL MODEL: POINT ESTIMATES
# =============================
# Obtain the expected number of renewals (point estimate).
        Renew <- fRenew(beta = Beta[i], eta = Eta[i], warranty = Wrnty, h = 1)

# Obtain the expected warranty cost.
        output$WrntWbl[i] <- Renew * CmpCst[i]

# Expected number of repairs
        output$NumRepWbl[i] <- Renew * Production

# WEIBULL MODEL: INTERVAL ESTIMATE - EXPECTED NUMBER OF RENEWALS AND WARRANTY COST
# ===========================================================================
# Obtain a confidence region for the Weibull parameters.
        degFree <- length(SurviveDF$VIN[SurviveDF$status == 1]) - 3

# Check the value of degFree
        if(degFree < 1)
            stop(message="Degrees of freedom < 1, since survival df has less than 4 points.")

        region <- fConfRgn(wblMdl, degFree, confLvl = 0.95, graph = F)
# If fConfRgn() returns "singular", wblMdl generated a singular matrix, so no inverse exits.
# Therefore, confidence intervals cannot be obtained. Should not get a singular matrix since
# the Weibull fit is adjusted above.
        if(region[[1]][[1]][1] != "singular")      # This extracts the first element
        {
# Obtain the expected number of renewals for these points.
            num <- length(region[[1]]$x)
            renRegn <- vector(mode = "numeric", length = num)
            renRegn <- as.numeric(rep(NA, num))

# For loop needed here as fRenew() requires its parameters as single values, not vectors.
            for(j in 1:num)
                renRegn[j] <- fRenew(beta = region[[1]]$x[j], eta = region[[1]]$y[j],
                    warranty = Wrnty, h = 4)

            RenewVar <- sum((renRegn-Renew)^2/length(renRegn[!is.na(renRegn)]), na.rm = T)
            RenewSD <- sqrt(RenewVar)
            output$WrntWblVar[i] <- Renew^2 * CmpCstSD[i]^2 +
              CmpCst[i]^2 * RenewVar + CmpCstSD[i]^2 * RenewVar
            output$WrntWblSD[i] <- sqrt(output$WrntWblVar[i])
        }
    }
  }
  detach("data")

# OUTPUT
# ======
  output
}
```

C.10 Sensitivitiy Analysis (*Sensitivity.ssc*)

```
#   WARRANTY COST SENSITIVITY ANALYSIS FOR A WEIBULL MODEL

#------------
# DESCRIPTION
#------------
# PURPOSE
# =======
# This function will calculate the percentage variation in warranty cost against the percentage
# variation in each of the Weibull parameters, both seperately and jointly. Graphs of the
# changes are produced.

# OUTLINE
# =======
#  For the inputted Part, an inputted cost data frame (CostDF) is searched to extract the
# corresponding values of beta and eta. A grid of percentage changes in beta and eta is then
# used to find the corresponding change in the warranty cost, which is graphed against each of
# the parameters seperately and then jointly. The warranty cost matrix is returned.

# INPUT
# =====
# 1. Part = The part id.
# 2. CostDF = The cost data frame, eg, costExpWbl.
# 3. Wrnty = The warranty length in years. Default = 3.
# 4. StepSize = Step size to be passed on to fRenew().

# OUTPUT
# ======
# The output is a matrix of the % change in warranty cost.

#---------
# PROGRAM
#---------
"fSenstv" <- function(Part, CostDF=costExpWbl, Wrnty=3, StepSize=4)
{
# SET UP VARIABLES
# ===============
# Extract the row number in CostDF that contains Part.
   rowNum <- as.numeric(row.names( CostDF[CostDF$Part==Part,] ))
# Set up grid to hold Beta and a warranty cost matrix.
   betaGrid <- seq(from=-0.1, to=0.1, by = 0.01)
   etaGrid <- seq(from=-0.1, to=0.1, by = 0.01)
   costGrid <- matrix(data=as.numeric(NA), nrow=length(betaGrid), ncol=length(etaGrid),
       dimnames=list(betaGrid,etaGrid))
   WrntCst <- CostDF$WrntWbl[rowNum]

# CALCULATE WARRANTY COST
# ======================
# Fill costGrid by calculating the warranty cost on every point (betaGrid, etaGrid).
   for (j in 1:length(etaGrid))
   {  for (i in 1:length(betaGrid))
       {
# Obtain the expected number of renewals and warranty cost.
          renew <- fRenew(beta=(1+betaGrid[i])*CostDF$Beta[rowNum],
              eta=(1+etaGrid[j])*CostDF$Eta[rowNum], warranty=Wrnty,
              h=StepSize)
          costGrid[i,j] <- renew * CostDF$CmpCst[rowNum]
       }
   }

# GENERATE GRAPHS
# ===============
# % change in cost against % change in beta.
   graphsheet()
   plot(betaGrid*100, (costGrid[,(length(etaGrid)+1)/2]-WrntCst)*100/WrntCst,
       xlab="% Change in Beta", ylab="% Change in Warranty Cost")
   title(CostDF$Name[CostDF$Part==Part])
```

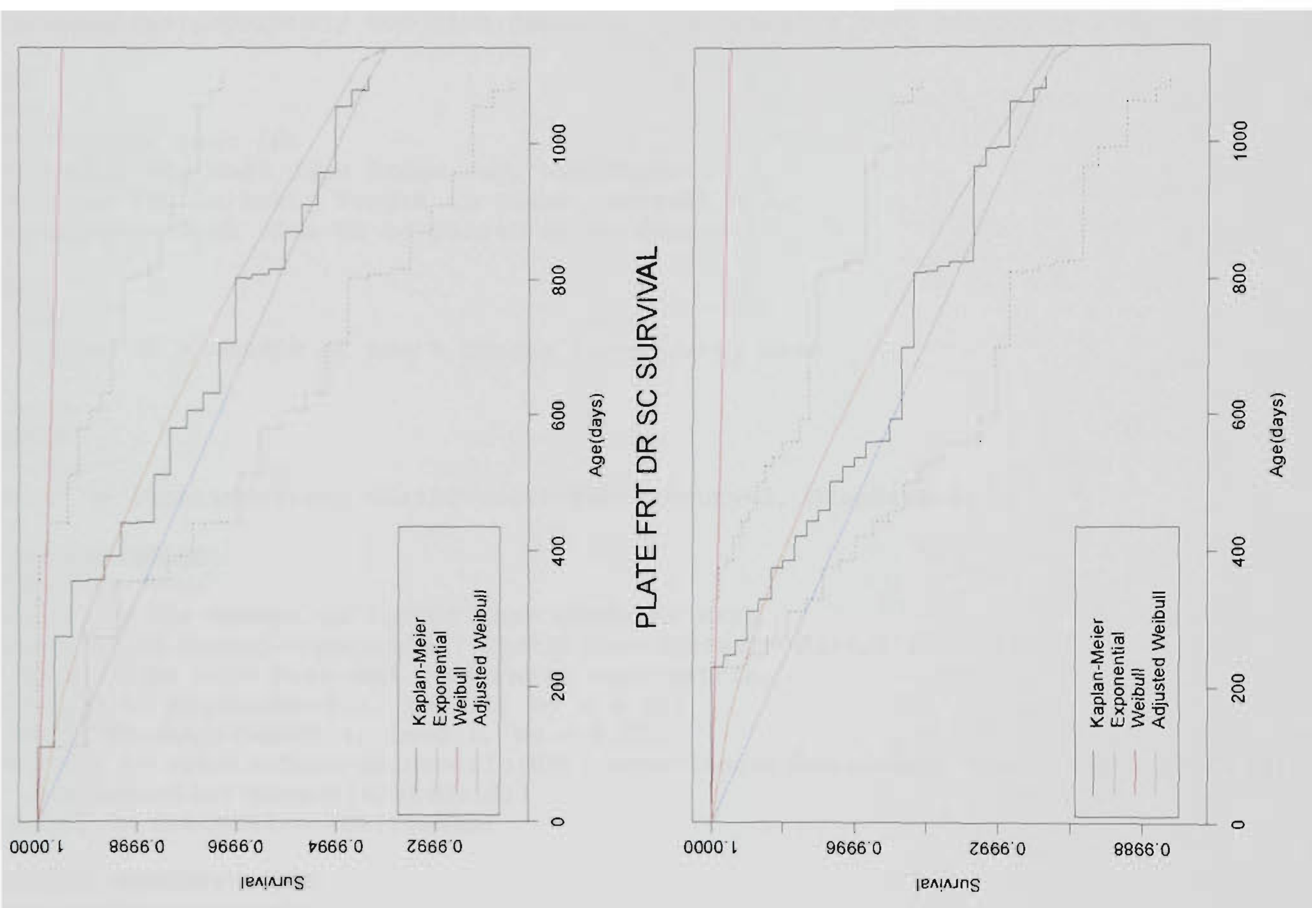
```
# % change in cost against % change in eta.
   plot(etaGrid*100, (costGrid[(length(betaGrid)+1)/2,]-WrntCst)*100/WrntCst,
      xlab="% Change in Eta", ylab="% Change in Warranty Cost")
   title(CostDF$Name[CostDF$Part==Part])

# % change in cost against % change in beta and eta (3-D plot).
# First calculate the range in cost.
   rangeCost <- (max(costGrid)-WrntCst)*100/WrntCst - (min(costGrid)-WrntCst)*100/WrntCst
   persp(betaGrid*100, etaGrid*100, (costGrid-WrntCst)*100/WrntCst,
      xlab="% Change in Beta",ylab="% Change in Eta", zlab="% Change in Warranty Cost" )
   title(CostDF$Name[CostDF$Part==Part])

# OUTPUT
# ======
# Return the cost grid.
   costGrid
}
```

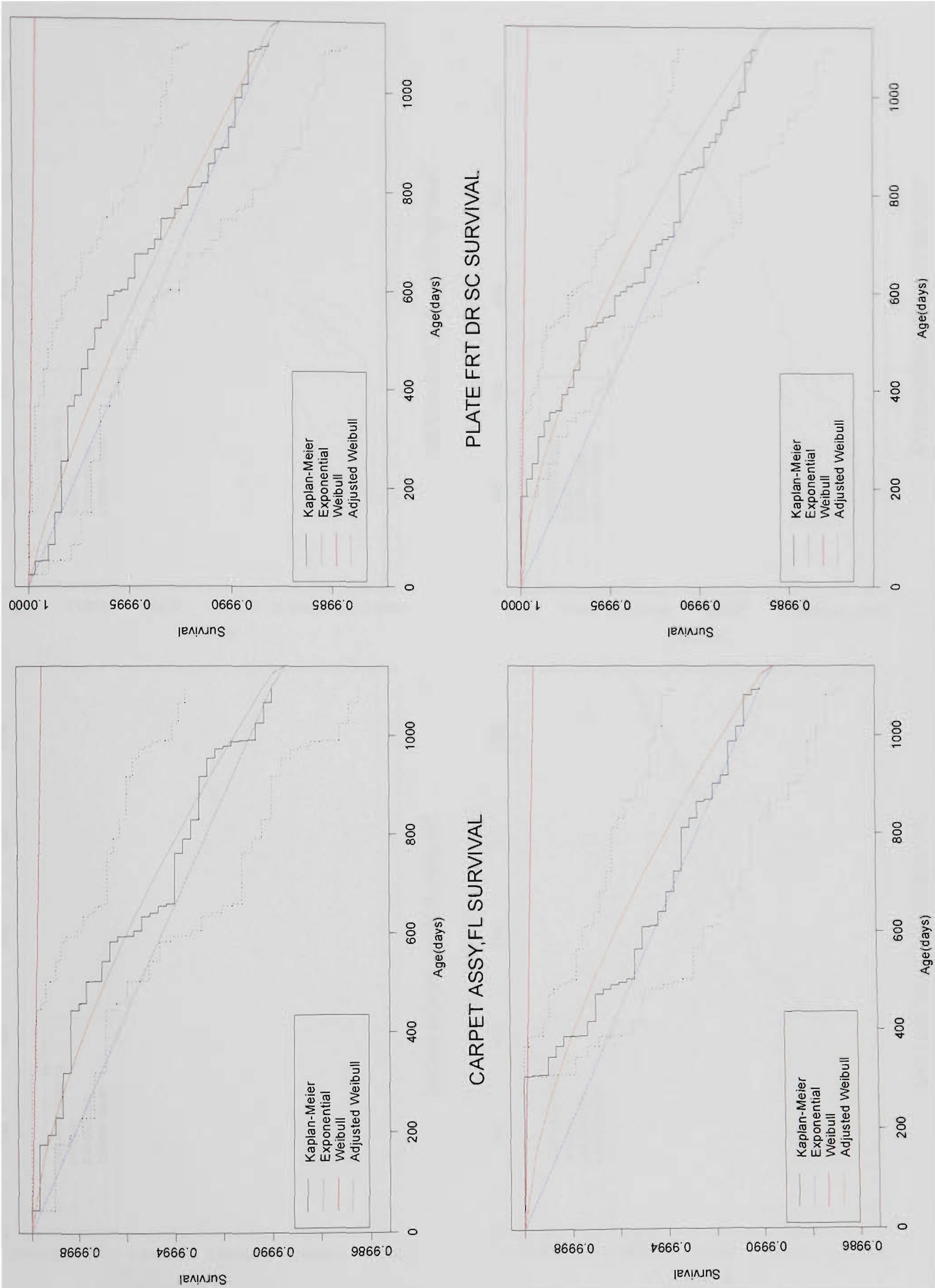
Appendix D: Graphs for Chapter 6

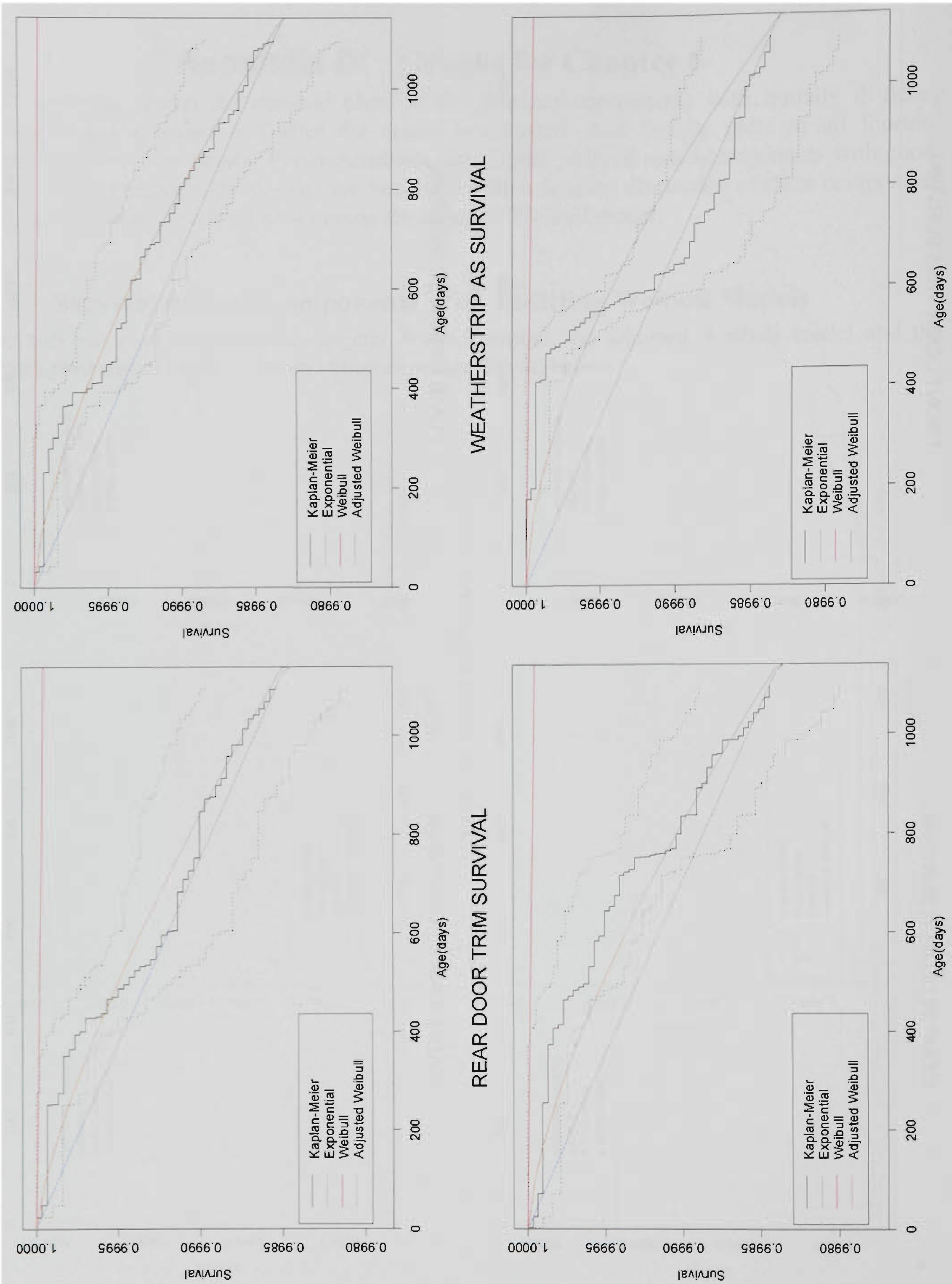
This appendix shows the survival plots of the fourteen components with initially ill-fitting Weibull models before and after the model is adjusted. Age density plots of all fourteen components are also shown. For comparison, age density plots of some components with good-fitting Weibull models are shown. See Section 6.3 for a detailed discussion of these components and the processing involved in obtaining the adjusted Weibull model.

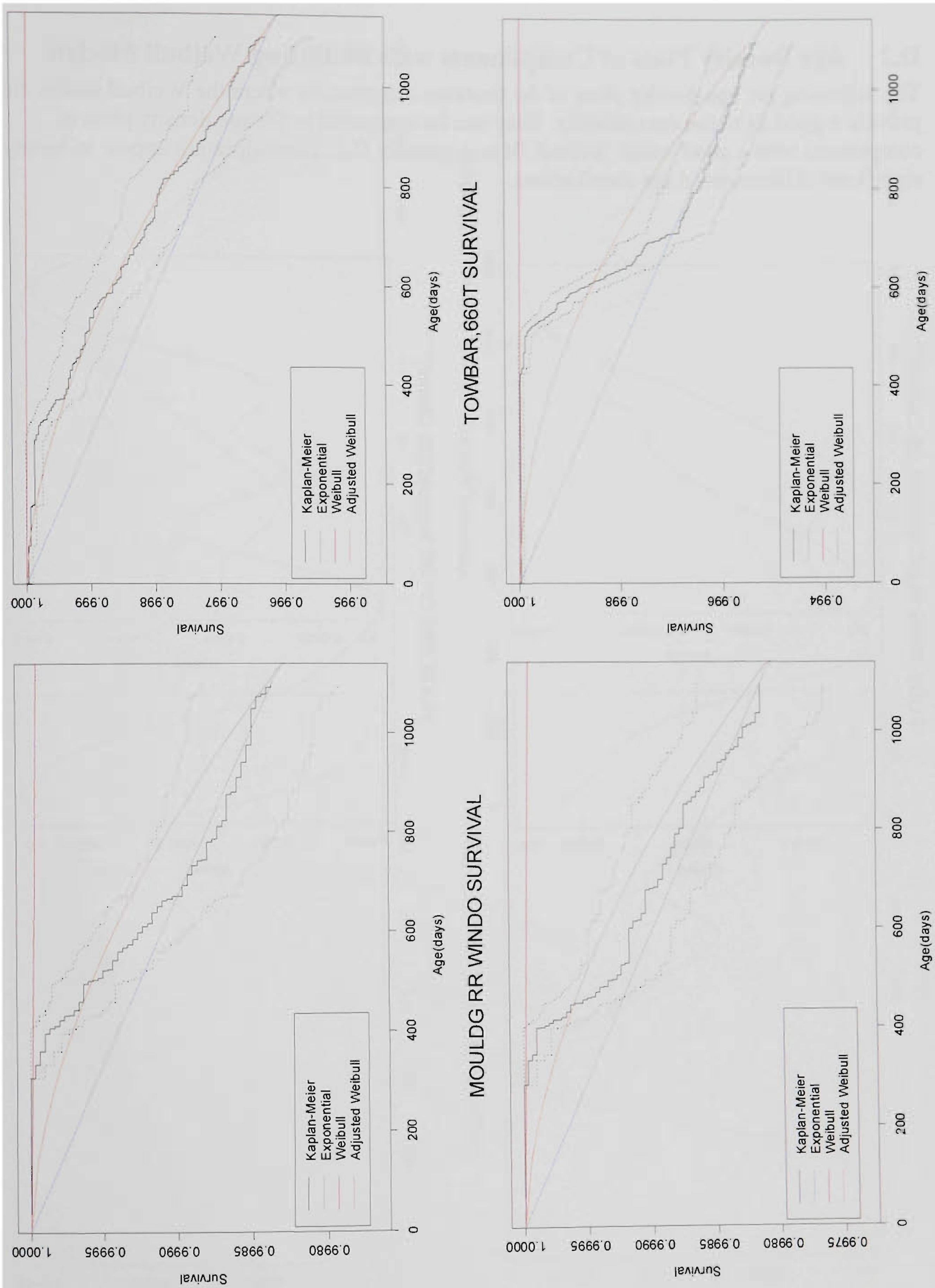
D.1 Survival Plots of Components With Ill-fitting Weibull Models

The survival plots showing the original Weibull model, the adjusted Weibull model and the exponential model of the fourteen components are shown below.



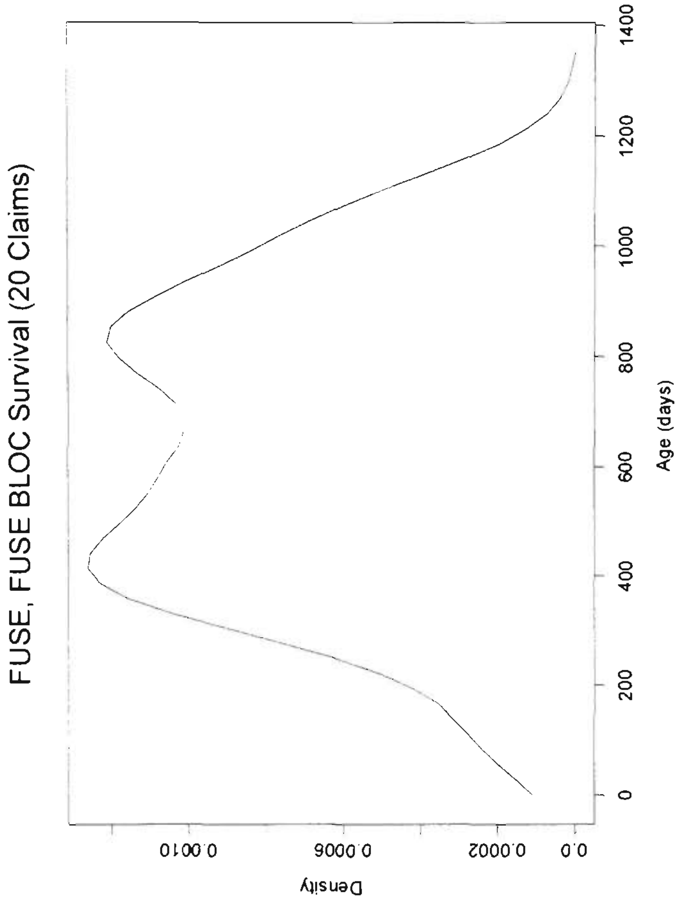

FRONT DOOR TRIM SURVIVAL
PLATE FRT DR SC SURVIVAL
COMP W CLUTCH SURVIVAL
CARPET ASSY,FL SURVIVAL
Kaplan-Meier
Exponential
Weibull
Adjusted Weibull
Age(days)
Survival

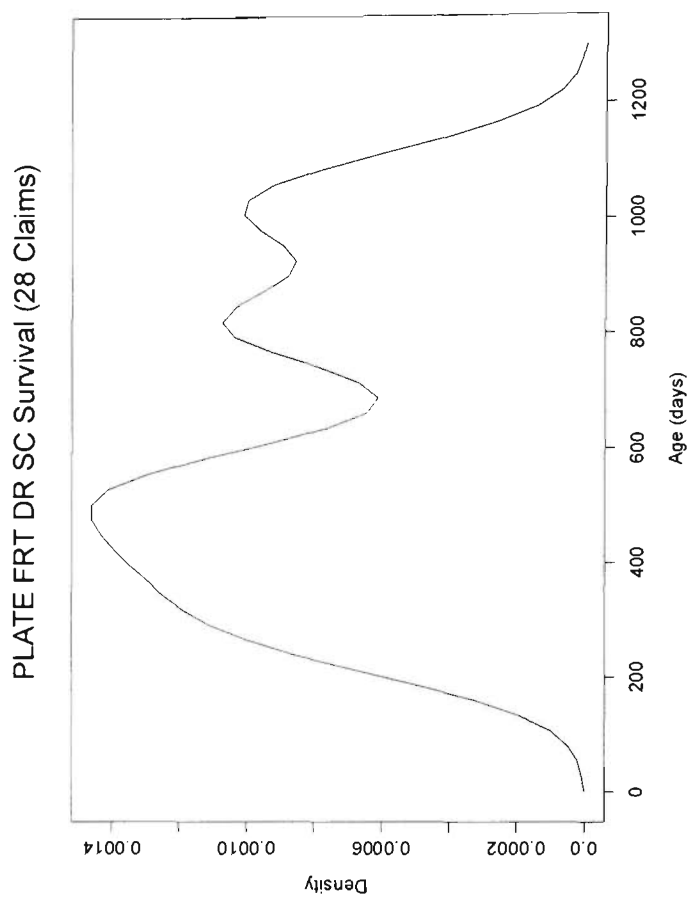

W/STRIP,RHF DOO SURVIVAL
WEATHERSTRIP AS SURVIVAL
W/STRIP,RFRL RH SURVIVAL
REAR DOOR TRIM SURVIVAL
Kaplan-Meier
Exponential
Weibull
Adjusted Weibull
Survival
Age(days)
1.0000
0.9995
0.9990
0.9985
0.9980
0
200
400
600
800
1000

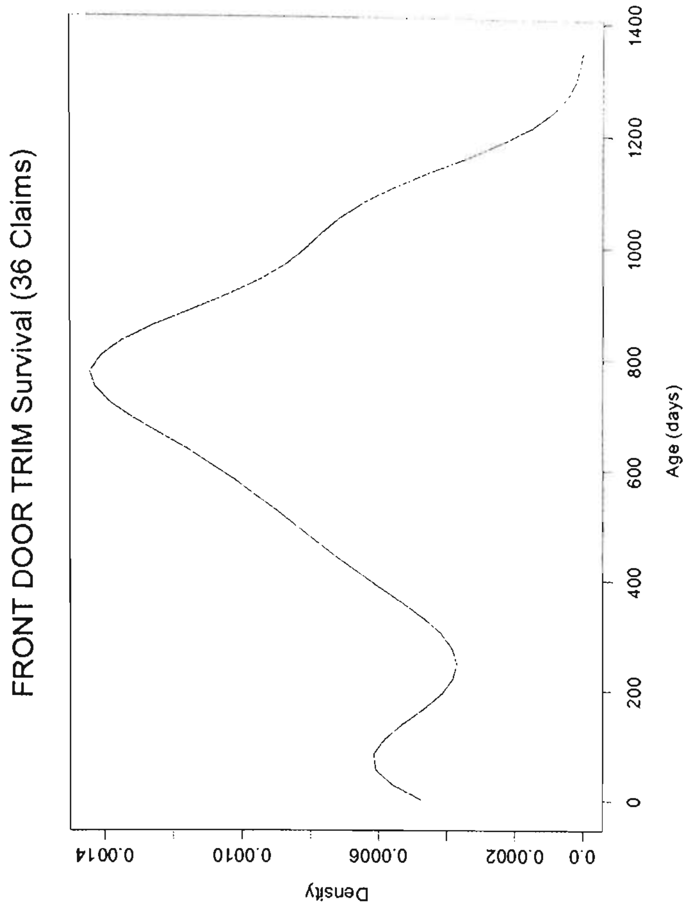

LOCK,RRDR LH V4 SURVIVAL
TOWBAR,660T SURVIVAL
LH RR MIRR ASSY SURVIVAL
MOULDG RR WINDO SURVIVAL
Kaplan-Meier
Exponential
Weibull
Adjusted Weibull
Survival
Age(days)
0
200
400
600
800
1000
1.000
0.999
0.998
0.997
0.996
0.995
0.994
1.0000
0.9995
0.9990
0.9985
0.9980
0.9975

D.2 Age Density Plots of Components with Ill-Fitting Weibull Models

The following are age density plots of the fourteen components where the Weibull model did not provide a good fit to the data initially. They can be compared to the age density plots of components with a good initial Weibull fit in Appendix D.3. There does not appear to be any significant differences in the distributions.






FRONT DOOR TRIM Survival (36 Claims)
Density
0.0 0.0002 0.0006 0.0010 0.0014
Age (days)
0 200 400 600 800 1000 1200 1400

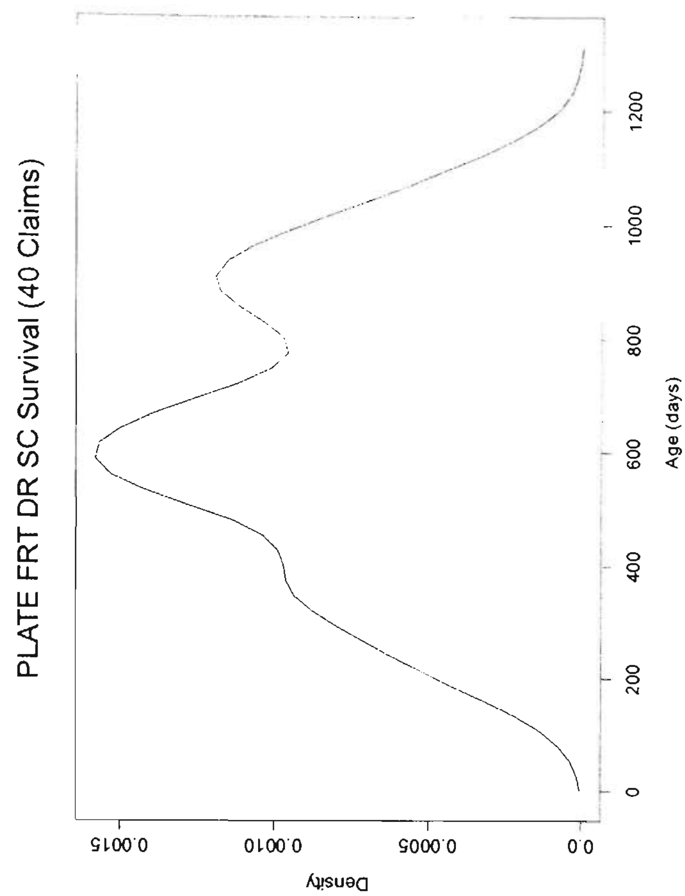
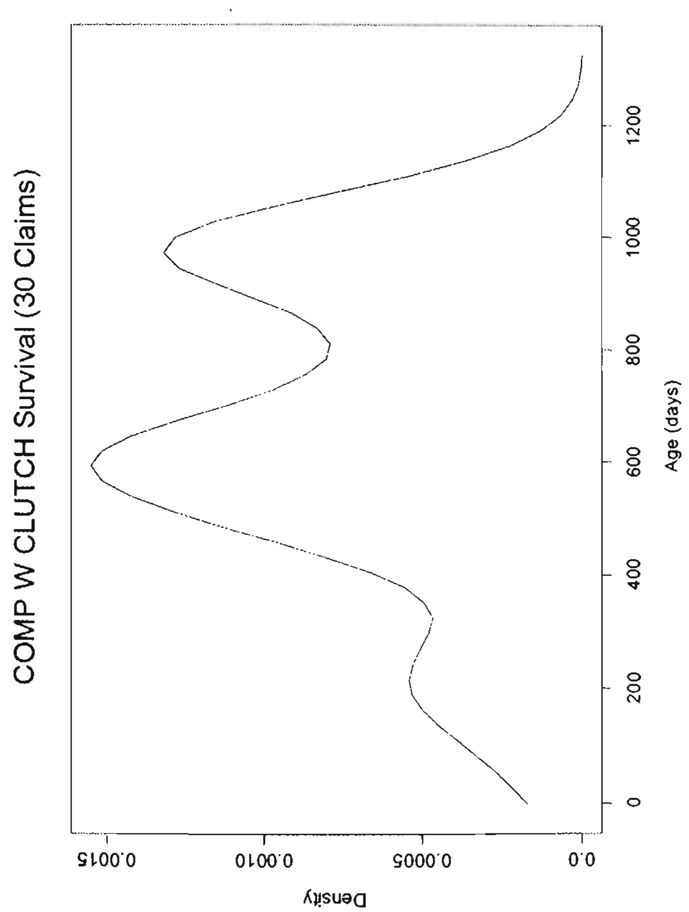
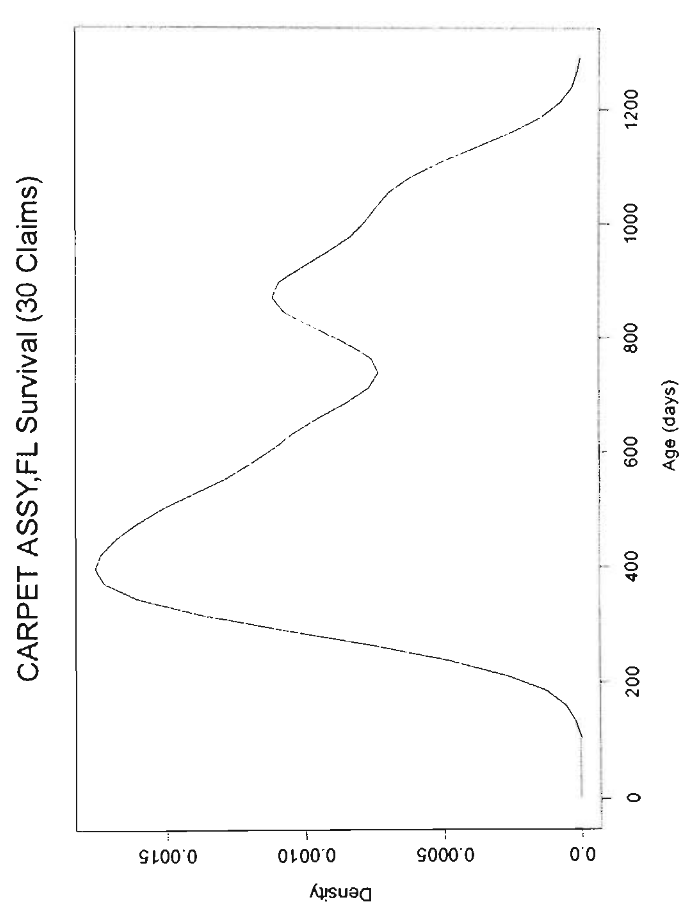
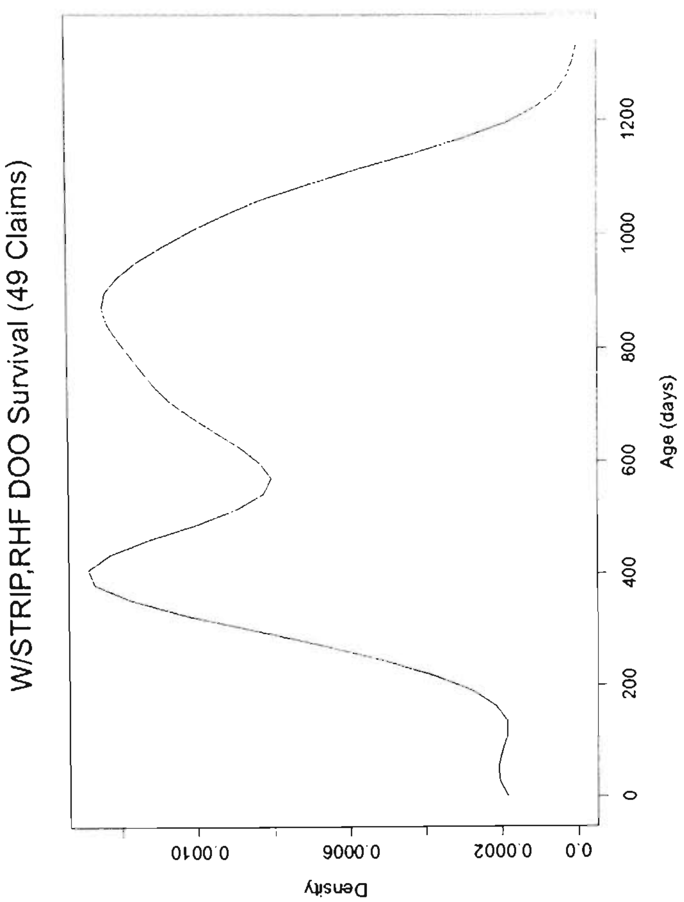
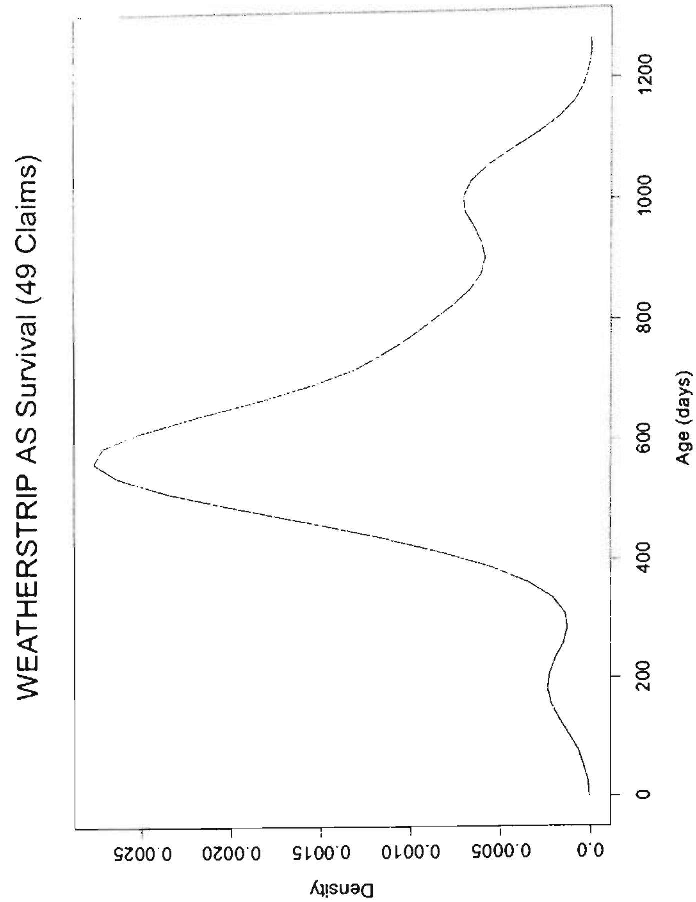


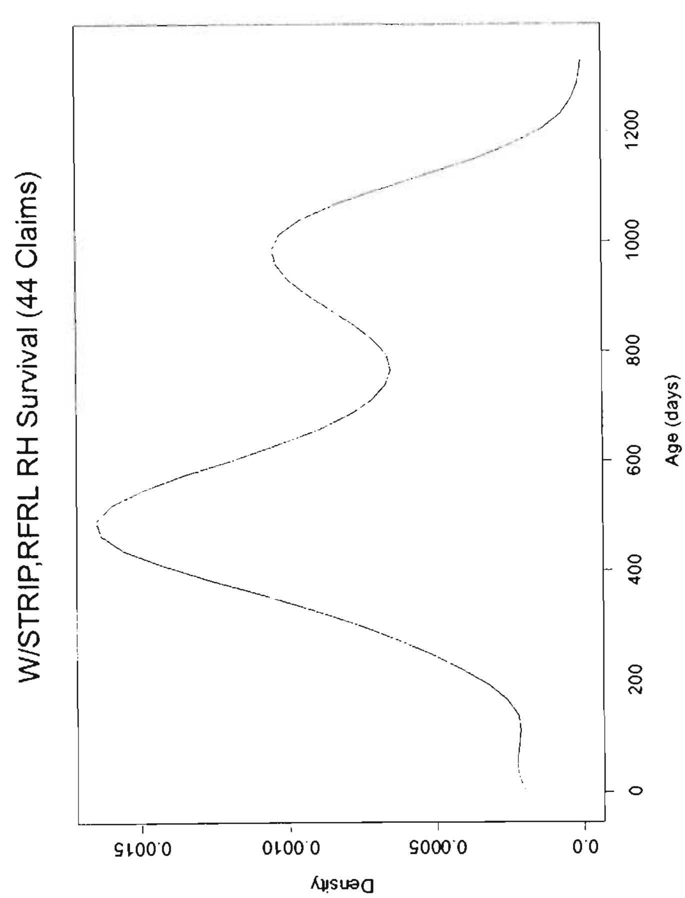
PLATE FRT DR SC Survival (40 Claims)
Density
0.0 0.0005 0.0010 0.0015
Age (days)
0 200 400 600 800 1000 1200

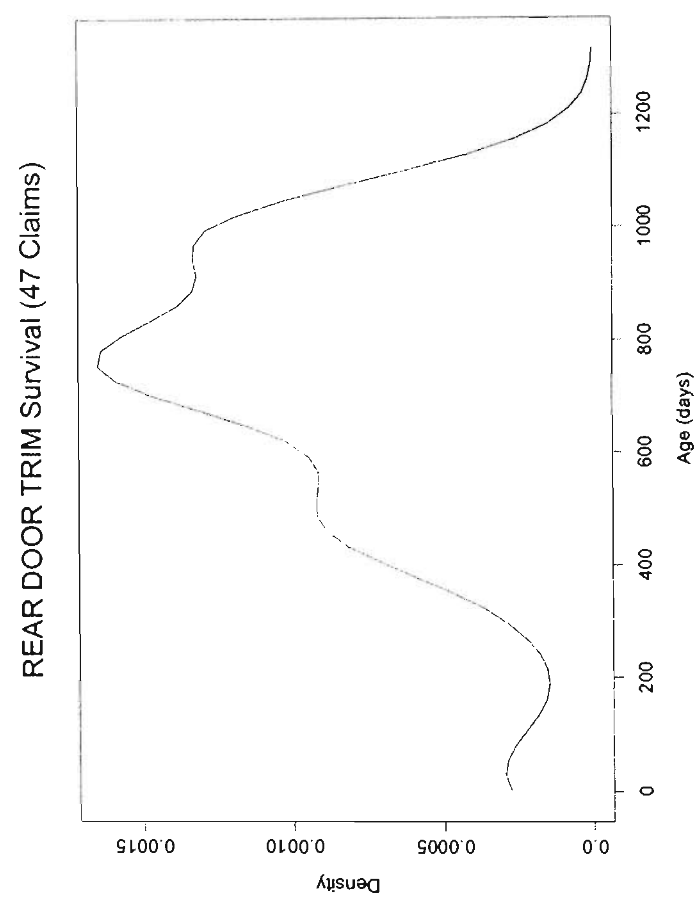

COMP W CLUTCH Survival (30 Claims)
Density
0.0 0.0005 0.0010 0.0015
Age (days)
0 200 400 600 800 1000 1200

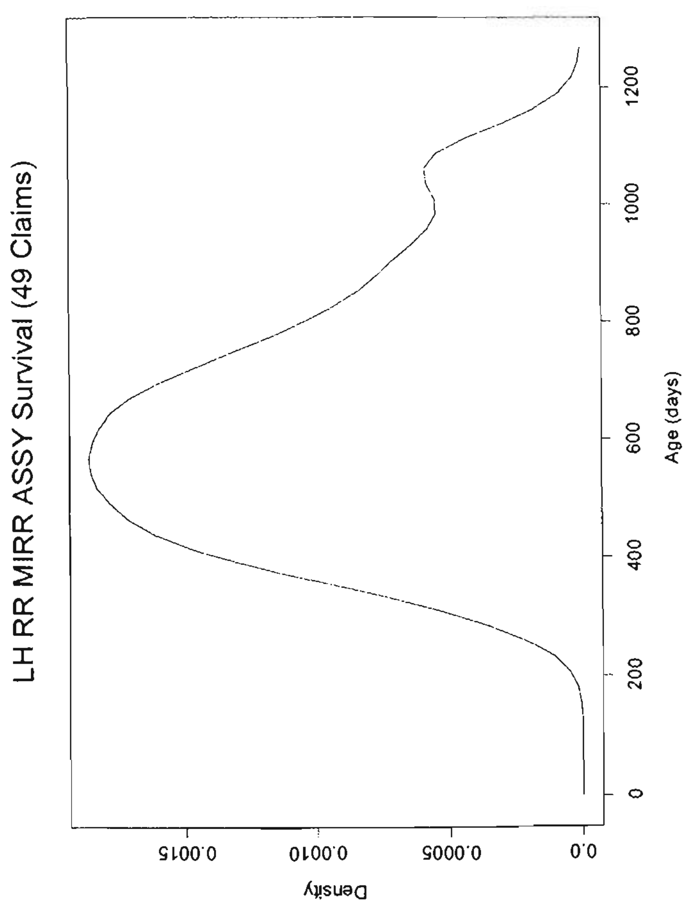

CARPET ASSY,FL Survival (30 Claims)
Density
0.0 0.0005 0.0010 0.0015
Age (days)
0 200 400 600 800 1000 1200

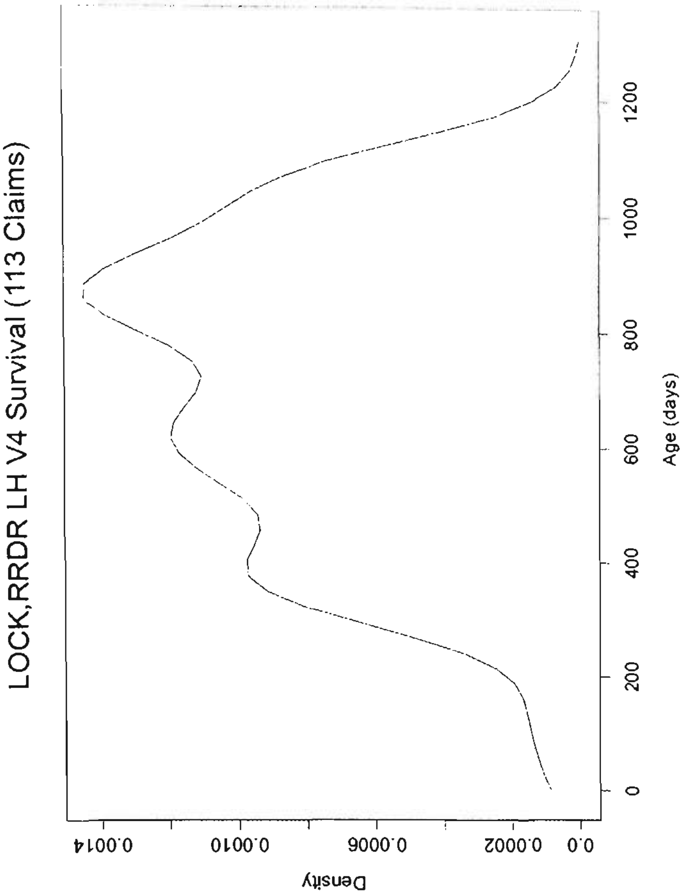

W/STRIP,RHF DOO Survival (49 Claims)
Density
0.0
0.0002
0.0006
0.0010
Age (days)
0
200
400
600
800
1000
1200

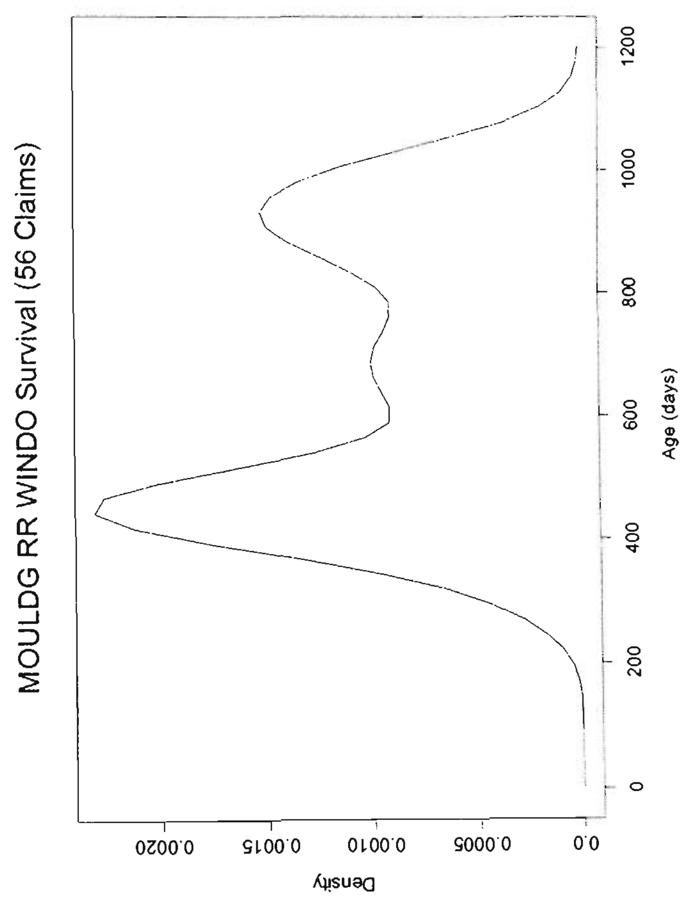

WEATHERSTRIP AS Survival (49 Claims)
Density
0.0
0.0005
0.0010
0.0015
0.0020
0.0025
Age (days)
0
200
400
600
800
1000
1200

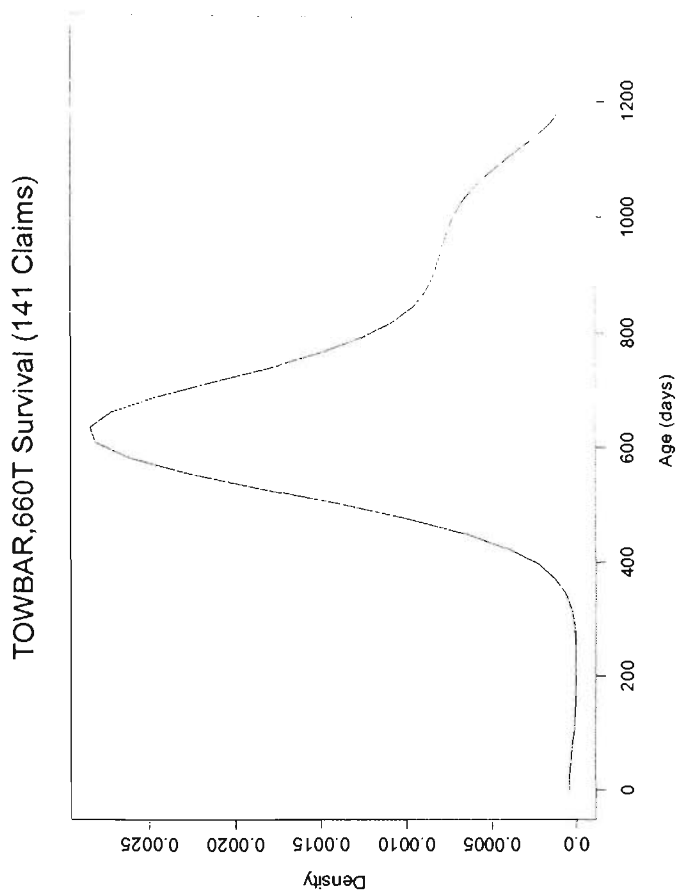

W/STRIP,RFRL RH Survival (44 Claims)
Density
0.0
0.0005
0.0010
0.0015
Age (days)
0
200
400
600
800
1000
1200


REAR DOOR TRIM Survival (47 Claims)
Density
0.0
0.0005
0.0010
0.0015
Age (days)
0
200
400
600
800
1000
1200


LH RR MIRR ASSY Survival (49 Claims)
Density
0.0
0.0005
0.0010
0.0015
Age (days)
0
200
400
600
800
1000
1200

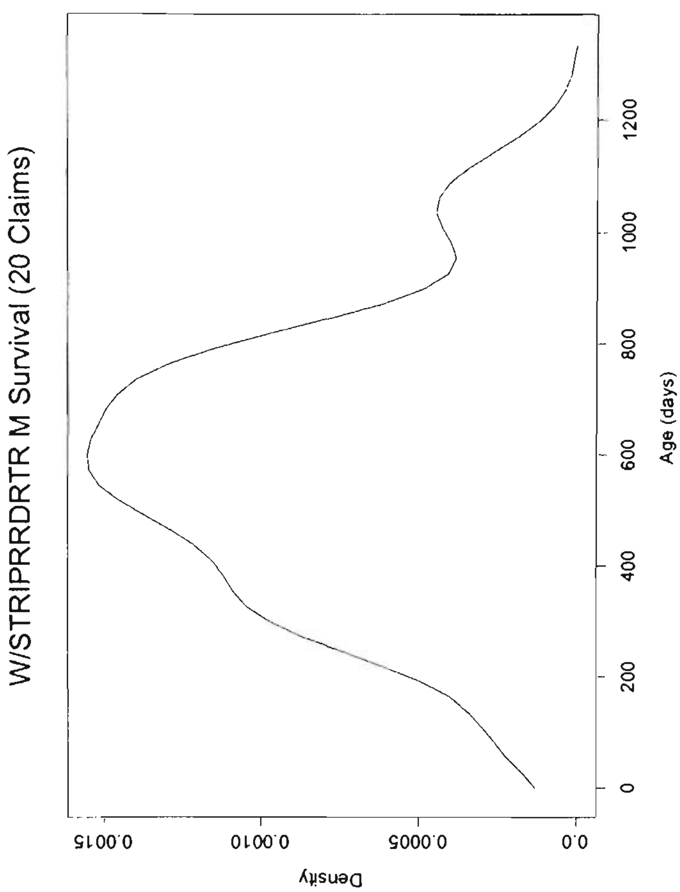

LOCK,RRDR LH V4 Survival (113 Claims)
Density
0.0
0.0002
0.0006
0.0010
0.0014
Age (days)
0
200
400
600
800
1000
1200

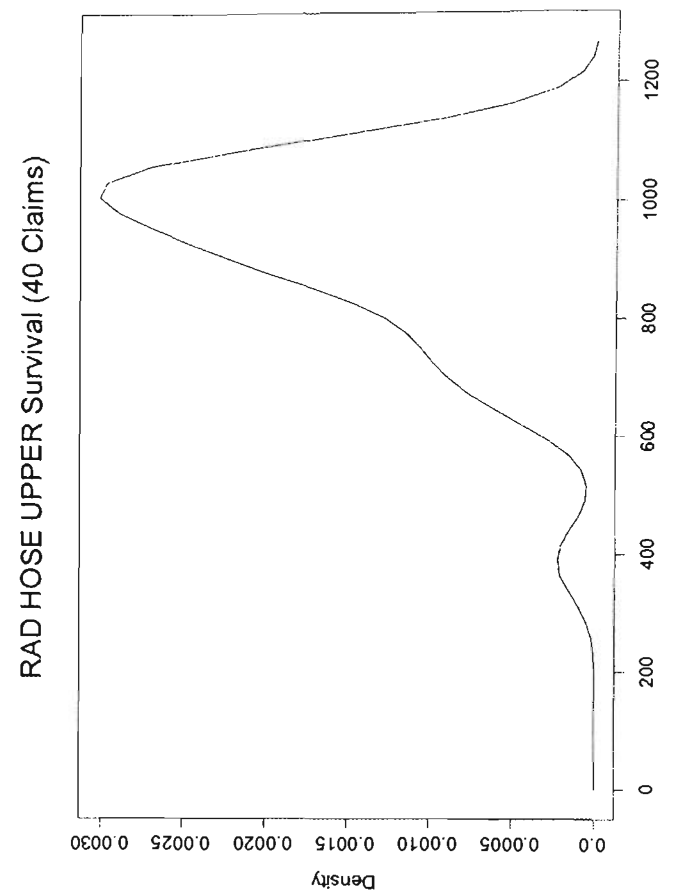

MOULDG RR WINDO Survival (56 Claims)
Density
0.0
0.0005
0.0010
0.0015
0.0020
Age (days)
0
200
400
600
800
1000
1200

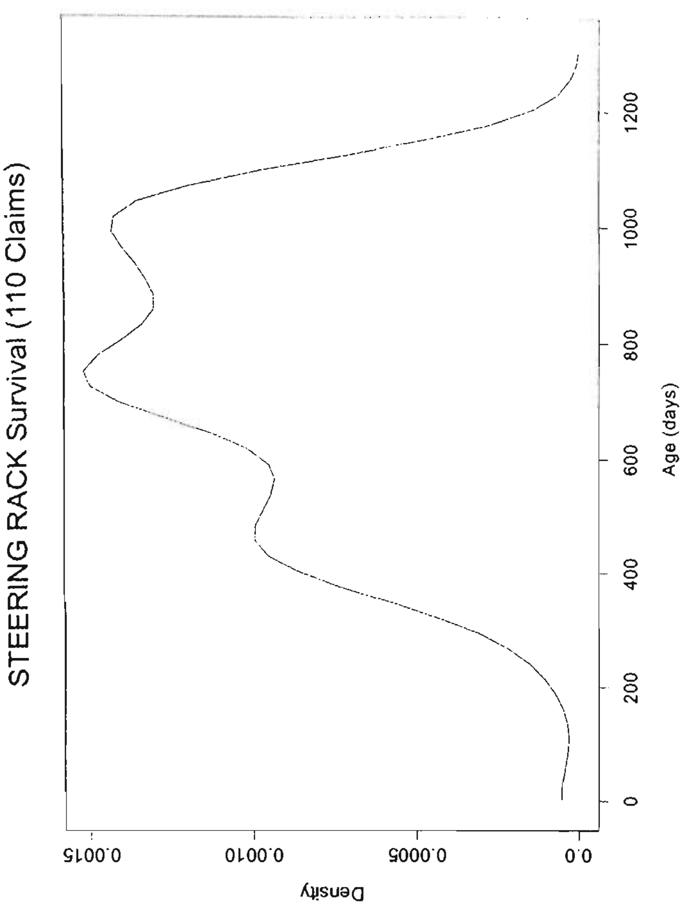

TOWBAR,660T Survival (141 Claims)
Density
0.0
0.0005
0.0010
0.0015
0.0020
0.0025
Age (days)
0
200
400
600
800
1000
1200

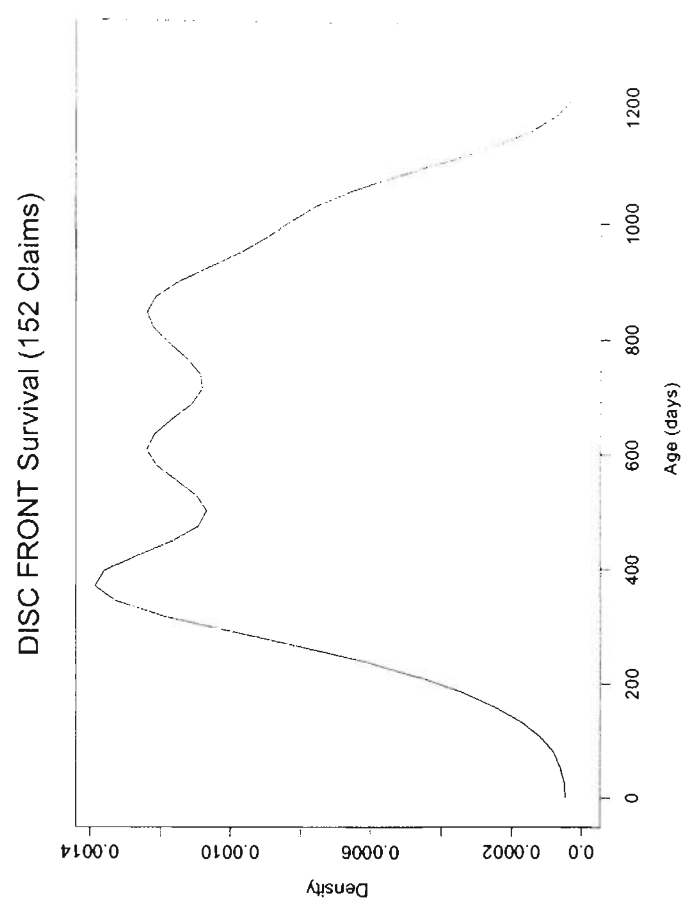
D.3 Age Density Plots of Components with Good-Fitting Weibull Models

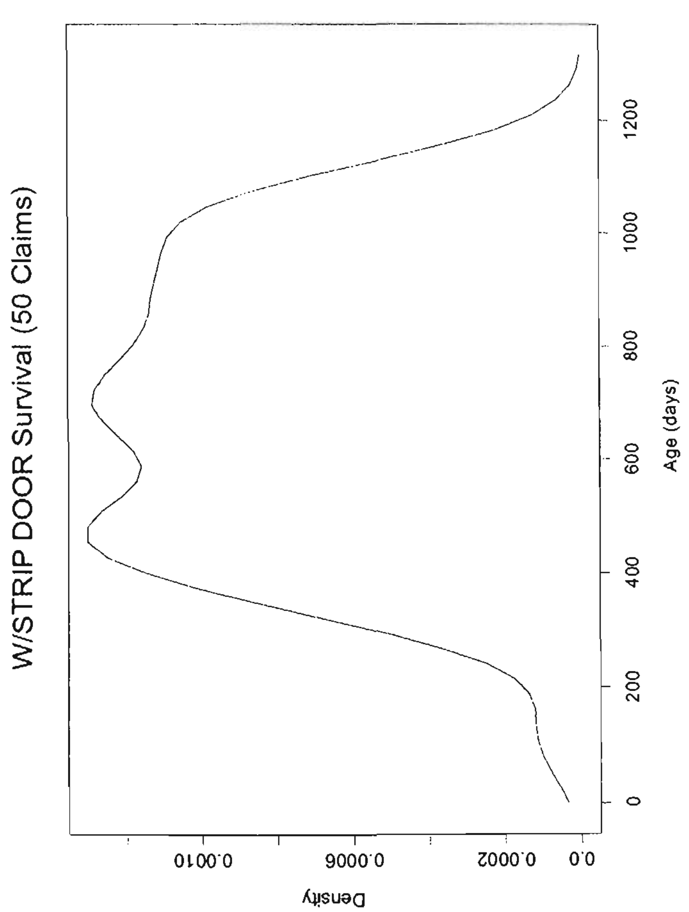
The following are age density plots of some components where the Weibull model provided a good fit to the data. There does not appear to be any significant difference between these plots and the age density plots of the plots of the components with an initially ill-fitting Weibull models shown in Appendix D.2.

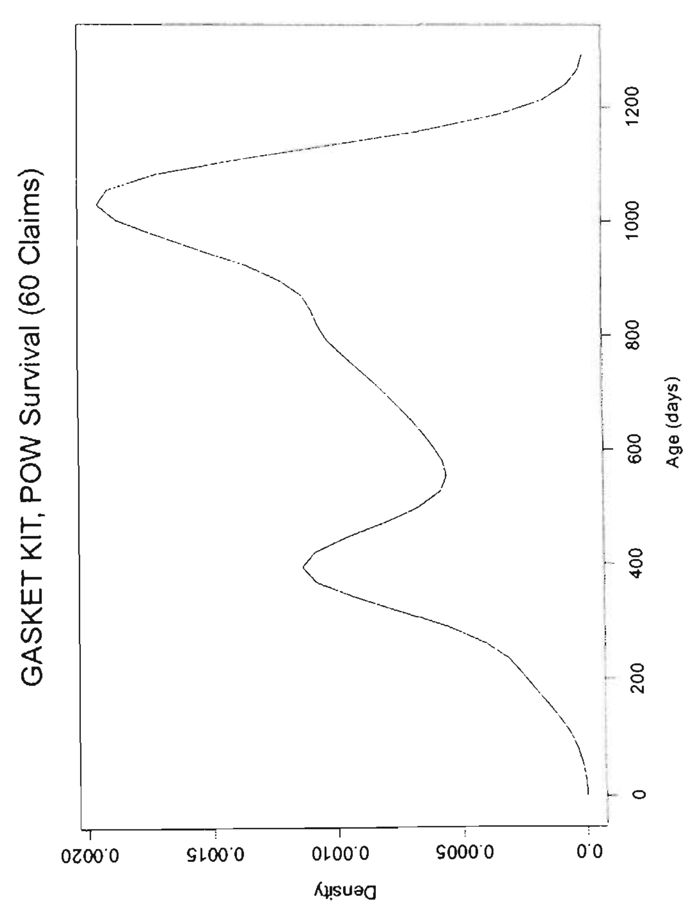





STEERING RACK Survival (110 Claims)
Density
Age (days)
0.0
0.0005
0.0010
0.0015
0
200
400
600
800
1000
1200


DISC FRONT Survival (152 Claims)
Density
Age (days)
0.0
0.0002
0.0006
0.0010
0.0014
0
200
400
600
800
1000
1200


W/STRIP DOOR Survival (50 Claims)
Density
Age (days)
0.0
0.0002
0.0006
0.0010
0
200
400
600
800
1000
1200


GASKET KIT, POW Survival (60 Claims)
Density
Age (days)
0.0
0.0005
0.0010
0.0015
0.0020
0
200
400
600
800
1000
1200

Appendix E: S-Plus Scripts for Chapter 7

This appendix contains scripts of functions that have been used in the simulations. The scripts are referred to in Chapter 7.

E.1 Warranty Cost Simulation (*WrntCstSimul.ssc*)

```
#      SIMULATED WARRANTY COST USING PARAMETRIC MODELLING

#------------
# DESCRIPTION
#------------
# PURPOSE
# =======
# This function creates a data frame of simulated warranty cost, or the warranty cost per
# vehicle.

# OUTLINE
# =======
# A random sample of size prod (inputted) is generated for each part using either the
# exponetial or Weibull model. Parts with wblClms (inputted) or more are modelled using the
# Weibull distribution. After generating the sample, the failure times are compared to wrnty
# (inputted) to identify which parts fail within the warranty. A sample of replaced parts is
# generated, which is again checked for failures within the warranty. This is repeated until
# the warranty runs out. The total warranty cost for each part and the number of replacements
# is returned.

# INPUT
# =====
# 1.   wc3yr = The 3-year warranty cost data frame. Default = costExpWbl.
# 2.   wc5yr = The 5-year warranty cost data frame. Default = extdWrnt
# 3.   compDF = The component data frame, containing the cost of repair. Default = compAge.
# 4.   wrnty = The warranty length in years. Default = 3 years.
# 5.   prod = The number of vehicles produced in a year. Default = 30138
# 6.   wblClms = The minimum number of claims for a part to be modelled by the Weibull
#      distribution. Default = 20.
# 7.   factor. If the 5-yr warranty cost is 'factor' times bigger than the 3-yr cost, then the
#      Weibull model is rejected in favour of the exponential. This is done because for some
#      parts, the cost of the 5-yr warranty is too many times bigger than the 3-yr.
# 8.   out = The ouput required. The two options are the warranty cost data frame ("df"), or
#      just the total warranty cost per vehicle ("cpv") (the default).

# OUTPUT
# ======
# The total warranty cost per vehcile of the simulated sample, or a data frame of warranty
# costs showing total warranty cost and total number of replacements, is returned.

#---------
# PROGRAM
#---------
"fWCstSim" <-
function(wc3yr=costExpWbl, wc5yr=extdWrnt, compDF=compAge, wrnty=3, prod=30138, wblClms=20,
   factor=5, out="cpv")
{

# CHECK INPUT PARAMETERS AND SET UP VARIABLES
# ===========================================
   if(wrnty <= 0)
      stop(message="The parameter 'Wrnty' (warranty period in years) should be positive.")
   if(prod <= 0 )
      stop(message="The parameter prod (vehicles produced in 1 year) should be positive.")
# Have to do this next one this way because the != doesn't seem to work properly
   if(out =="cpv" | out == "df") T
#  if(out != "cpv" | out != "df")
   else
   warning("The parameter 'out' must be either 'df' (data frame) or 'cpv' (cost per vehicle)")
```

```
# Convert warranty to days.
  WrntDays <- round(wrnty * 365.25)

# Set up data frame for warranty costs.
  wc <- data.frame(Part=NA,Cost=0,Rep=0)

# CHECK HOW MANY TIMES BIGGER THE 5-YR WARRANTY COST IS COMPARED TO THE 3-YR
# ======================================================================
# When the 5-yr warranty cost is more than 'factor' times bigger than the 3-yr cost under the
# Weibull model, use the exponential model instead.
  cond <- wc3yr$Claims >= wblClms & wc3yr$Part != ""
  extWr2big <- wc3yr$Part[cond][wc5yr$WrntWbl[cond] >
      factor * wc3yr$WrntWbl[cond] & !is.na(wc5yr$WrntWbl[cond])]

# GENERATE FAILURE TIMES AND WORK OUT WARRANTY COST FOR EACH PART
# ==============================================================
  for( i in 1:length(wc3yr$Part) )
  {
# Initialise the total number of times current part is replaced in all production to 0.
  totRep <- 0

# Counter to be used as an escape clause if while loop jams.
  counter <- 0

# Set failure times using appropriate model. Note: Failure times of parts with no id number
# are simulated using the exponential model with a rate covering all the parts. If it's a 5-yr
# warranty simulation and the 5-yr cost was more than 'factor' times the 3-yr cost, use the
# exponential model.
     if( wc3yr$Claims[i] < wblClms | wc3yr$Part[i] =="" |
     (wrnty==5 & is.element(wc3yr$Part[i], extWr2big)) )
        carAge <- rexp(n=prod, rate=wc3yr$Rate[i])
     else
        carAge <- rweibull(n=prod, shape=wc3yr$Beta[i], scale=wc3yr$Eta[i])

# See which and how many vehicles fail on current part during warranty.
        inWrnty <- carAge <= WrntDays
        numRep <- length(inWrnty[inWrnty==T])

# The while loop will be entered every time current part fails in a vehicle until warranty
  ends.
     while(numRep > 0)
     {
# Add the number of repairs to the total.
        totRep <- totRep + numRep

# Set vehicle age as sum of its previous age and time replacement part fails.
        if( wc3yr$Claims[i] < wblClms | wc3yr$Part[i] =="" |
        (wrnty==5 & is.element(wc3yr$Part[i], extWr2big)) )
           carAge[inWrnty] <- carAge[inWrnty] + rexp(n=numRep, rate=wc3yr$Rate[i])
        else
           carAge[inWrnty] <- carAge[inWrnty] +
              rweibull(n=numRep, shape=wc3yr$Beta[i], scale=wc3yr$Eta[i])

# See which and how many vehicles with replaced parts fail again during warranty.
        inWrnty <- carAge <= WrntDays
        numRep <- length(inWrnty[inWrnty==T])

# Increment counter and check its value.
        counter <- counter + 1
        if(counter >= 50)
        {  warning("While loop terminated after 50 part replacements on one vehicle.")
           break
        }
     }

# Randomly choose the cost of repair from the component repair cost in compDF. An alternative
# would be to use a normal approximation to the distribution of the cost using the mean cost
```

```
# and sd in wc3yr, but the cost is not quite normally distributed for most parts.
      partWrtCst <-
         sum( sample(compDF$cost[compDF$ofp==wc3yr$Part[i]], size=totRep, replace=T) )

      wc[i,"Part"] <- wc3yr$Part[i]
      wc[i,"Cost"] <- partWrtCst
      wc[i,"Rep"] <- totRep
   }

# OUTPUT
# ======
# Return either the warranty cost data frame (wc) or only the sum of the warranty cost,
# depending on what was chosen in the input parameter.
   if(out=="df")
      wc
   if(out=="cpv")
      sum(wc$Cost)/prod
}
```

E.2 Simulated Warranty Cost With Distance Limitation (*WrntCstSmulDist.ssc*)

```
#     SIMULATED WARRANTY COST WITH DISTANCE LIMITATION, USING PARAMETRIC MODELLING

#------------
# DESCRIPTION
#------------
# PURPOSE
# =======
# This function creates a data frame of simulated warranty cost, or the warranty cost per
# vehicle.

# OUTLINE
# =======
# A random sample of size prod (inputted) is generated for each part using either the
# exponential or Weibull model. Parts with wblClms (inputted) or more are modelled using the
# Weibull distribution. After generating the sample, the failure times are compared to wrnty
# (inputted) to identify which parts fail within the warranty. A sample of replaced parts is
# generated, which is again checked for failures within the warranty. This is repeated until
# the warranty runs out. The total warranty cost for each part and the number of replacements
# is returned.

# INPUT
# =====
# 1.   wc3yr = The warranty cost data frame. Default = costExpWbl.
# 2.   compDF = The component data frame, containing the cost of repair. Default = compAge.
# 3.   wrnty = The warranty length in years. Default = 3 years.
# 4.   prod = The number of vehicles produced in a year. Default = 30138
# 5.   wblClms = The minimum number of claims for a part to be modelled by the Weibull
#      distribution. Default = 20.
# 6.   factor. If the 5-yr warranty cost is 'factor' times bigger than the 3-yr cost, then the
#      Weibull model is rejected in favour of the exponential. This is done because for some
#      parts, the cost of the 5-yr warranty is too many times bigger than the 3-yr.
# 7.   out = The ouput required. The two options are the warranty cost data frame ("df"), or
#      just the total warranty cost per vehicle ("cpv") (the default).

# OUTPUT
# ======
# The total warranty cost per vehcile of the simulated sample, or a data frame of warranty
# costs showing total warranty cost and total number of replacements, is returned.

#---------
# PROGRAM
#---------
"fWCstSmD" <-
function(wc3yr=costExpWbl, wc5yr=extdWrnt, compDF=compAge, wrnty=3, wrntDist=100000,
   prod=30138, wblClms=20, factor=5, out="cpv")
{
```

```
# CHECK INPUT PARAMETERS AND SET UP VARIABLES
# ==========================================
   if(wrnty <= 0)
      stop(message="The parameter 'Wrnty' (warranty period in years) should be positive.")
   if(prod <= 0 )
      stop(message="The parameter prod (vehicles produced in 1 year) should be positive.")
# Have to do this next one this way because the != doesn't seem to work properly
   if(out =="cpv" | out == "df") T
#  if(out != "cpv" | out != "df")
   else
   warning("The parameter 'out' must be either 'df' (data frame) or 'cpv' (cost per vehicle)")

# Convert warranty to days.
   WrntDays <- round(wrnty * 365.25)

# Set up data frame for warranty costs.
   wc <- data.frame(Part=NA,Cost=0,Rep=0)

# CHECK HOW MANY TIMES BIGGER THE 5-YR WARRANTY COST IS COMPARED TO THE 3-YR
# ==========================================================================
# When the 5-yr warranty cost is more than 'factor' times bigger than the 3-yr cost under the
# Weibull model, use the exponential model instead.
   cond <- wc3yr$Claims >= wblClms & wc3yr$Part != ""
   extWr2big <- wc3yr$Part[cond][wc5yr$WrntWbl[cond] >
      factor * wc3yr$WrntWbl[cond] & !is.na(wc5yr$WrntWbl[cond])]

# GENERATE FAILURE TIMES AND WORK OUT WARRANTY COST FOR EACH PART
# ===============================================================
   for( i in 1:length(wc3yr$Part) )
   {
# Initialise the total number of times current part is replaced in all production to 0.
   totRep <- 0

# Counter to be used as an escape clause if while loop jams.
   counter <- 0

# Set failure times using appropriate model. Note: Failure times of parts with no id number
# are simulated using the exponential model with a rate covering all the parts. If it's a 5-yr
# warranty simulation and the 5-yr cost was more than 'factor' times the 3-yr cost, use the
# exponential model.
      if( wc3yr$Claims[i] < wblClms | wc3yr$Part[i] =="" |
      (wrnty==5 & is.element(wc3yr$Part[i], extWr2big)) )
         carAge <- rexp(n=prod, rate=wc3yr$Rate[i])
      else
         carAge <- rweibull(n=prod, shape=wc3yr$Beta[i], scale=wc3yr$Eta[i])

# Set usage rate.
      useRate <- sample(usage, size=prod, replace=T)

# See which and how many vehicles fail on current part during warranty.
      inWrnty <- carAge <= WrntDays & useRate * carAge/365.25
      numRep <- length(inWrnty[inWrnty==T])

# The while loop will be entered every time current part fails in a vehicle until warranty
  ends.
      while(numRep > 0)
      {
# Add the number of repairs to the total.
         totRep <- totRep + numRep

# Set vehicle age as sum of its previous age and time replacement part fails.
         if( wc3yr$Claims[i] < wblClms | wc3yr$Part[i] =="" |
         (wrnty==5 & is.element(wc3yr$Part[i], extWr2big)) )
            carAge[inWrnty] <- carAge[inWrnty] + rexp(n=numRep, rate=wc3yr$Rate[i])
         else
            carAge[inWrnty] <- carAge[inWrnty] +
            rweibull(n=numRep, shape=wc3yr$Beta[i], scale=wc3yr$Eta[i])
```

```
# See which and how many vehicles with replaced parts fail again during warranty.
          inWrnty <- carAge <= WrntDays & useRate * carAge/365.25
          numRep <- length(inWrnty[inWrnty==T])

# Increment counter and check its value.
          counter <- counter + 1
          if(counter >= 50)
          {   warning("While loop terminated after 50 part replacements on one vehicle.")
              break
          }
       }

# Randomly choose the cost of repair from the component repair cost in compDF. An alternative
# would be to use a normal approximation to the distribution of the cost using the mean cost
# and sd in wc3yr, but the cost is not quite normally distributed for most parts.
       partWrtCst <-
          sum( sample(compDF$cost[compDF$ofp==wc3yr$Part[i]], size=totRep, replace=T) )

       wc[i,"Part"] <- wc3yr$Part[i]
       wc[i,"Cost"] <- partWrtCst
       wc[i,"Rep"] <- totRep
    }

# OUTPUT
# ======
# Return either the warranty cost data frame (wc) or only the sum of the warranty cost,
# depending on what was chosen in the input parameter.
   if(out=="df")
       wc
   if(out=="cpv")
       sum(wc$Cost)/prod
}
```

E.3 Sample Generating Function (*SmpDisc.ssc*)

```
#          RANDOM SAMPLE GENERATION OF ONE YEAR'S PRODUCTION

#------------
# DESCRIPTION
#------------
# PURPOSE
# =======
# This function creates a sample with given Weibull parameters.

# OUTLINE
# =======
# Failure times for a sample of the inputted sample size (one year's production) are generated
# using the inputted Weibull shape and scale parameters. If the failure times are less than
  the
# inputted warranty length, new failure times are generated, until the warranty on all of the
# sample expires. From the simulated failure times, a Weibull model is fitted and the
# parameters returned. This can then be repeated many times to obtain a distribution of fitted
# Weibull parameters.

# INPUT
# =====
# 1.   sampSize = the size of the generated sample, usually set to one year's production.
# 2.   wrnty =  the warranty length in years.
# 3.   beta = the shape parameter of the Weibull distribution from which a sample is
  generated.
# 4.   eta = the scale parameter of the Weibull distribution from which a sample is generated.

# OUTPUT
# ======
# The values of the Weibull paramters of the fitted model to the generated data.
```

```
#---------
# PROGRAM
#---------
"fSmpPrms" <-
function(sampSize, wrnty, beta, eta)
{
# Convert wrnty from years to days.
   wrnty <- round(365.25*wrnty)
   samp <- round( rweibull(sampSize, shape=beta, scale=eta) )

# Create a data frame with fields: ID, component's age at failure (compAge), failure/censored
# status (fail=1), vehicle's age at failure (vehAge) and component's age (age), be it censored
# or observed, at the end of the warranty. The field "age" will be set to wrnty, and adjusted
# to a part's age if it fails before the warratny ends.
   sampleDF <- data.frame(cbind(ID=1:sampSize, compAge=samp, status=0, vehAge=samp,
       age=wrnty))

# Identify records where failure occurs within the warranty period and count how many there
   are.
   inWrnty <- sampleDF$vehAge < wrnty
   numFailed <- count.rows(sampleDF[inWrnty,])

# Set status to 1 if part fails before end of warrranty.
   sampleDF$status[inWrnty] <- 1

# Set age to compAge if part fails within warranty and wrnty if part has not failed.
   sampleDF$age[inWrnty] <- sampleDF$compAge[inWrnty]

# While there are vehicles in warranty, replace part, and add a new record.
   count <- 1
   while(numFailed > 0)
   {
       newSamp <- rweibull(numFailed, shape=beta, scale=eta)
       newRecords <- data.frame(cbind(ID=sampleDF$ID[inWrnty],
           compAge=newSamp, status=0, vehAge=sampleDF$vehAge[inWrnty]+newSamp,
           age=wrnty-sampleDF$vehAge[inWrnty]) )

# Exclude records of parts that have been replaced from generating new records by changing
# vehAge to 99999, which is beyond  the warrnty. The vehicles's age will be kept by the
# replaced part record.
       sampleDF$vehAge[inWrnty] <- 99999

# Add the new records to the existing ones.
       sampleDF <- rbind(sampleDF, newRecords)

# See which vehicles are in warranty in the new combined data frame.
       inWrnty <- sampleDF$vehAge < wrnty

# Set status to 1 if part fails before end of warrranty.
       sampleDF$status[inWrnty] <- 1

# Set age to compAge if part fails within warranty and wrnty if part has not failed.
       sampleDF$age[inWrnty] <- sampleDF$compAge[inWrnty]

# See how many are still in warranty.
       numFailed <- count.rows(sampleDF[inWrnty,])

# Escape clause if there is a problem.
       count <- count + 1
       if(count==50)
           stop(message="Error: Generated 50 new samples already.")
   }

# Fit a Weibull distribution to the sample.
       WblMdl <- survReg(Surv(age + 1, status, type = "right") ~ 1, data = sampleDF,
           na.action = na.exclude, dist = "weibull", scale=0,
           control=list(maxiter = 100, rel.tolerance=1e-004, failure=1))
```

```
# Return the Weibull parameters. The shape parameter (beta) is 1/WblMdl$scale and the scale
# parameter (eta) is exp(WblMdl$coeff).
   matrix(c(1/WblMdl$scale, exp(WblMdl$coeff) ), nrow=1, ncol=2)
}
```

E.4 Production Density (*ProdDens.ssc*)

```
#       PRODUCTION DENSITY

#-------------
# DESCRIPTION
#------------
# INTRODUCTION
# ============
# This function creates boxplots and density plots of vehicle production throughtout the year
# if the input Plot parameter is set to T.

# OUTPUT
# ======
# A vector of the weekly median production using the sales 97 and 99 data frames.

#---------
# PROGRAM
#---------
"fProdDns" <-
function (Plot=T)
{
# Calculate the day of the year that vehicles were produced in 97 and 99.
   prod97 <- sales97$prod - "12/31/1996"
   prod99 <- sales99$prod - "12/31/1998"

# Convert these to densities.
   dens97 <- density(prod97, n=365, from=1, to=365)
   dens99 <- density(prod99, n=365, from=1, to=365)

# Combine these two into one "average" production and obtain its density.
   prod <- c(prod97, prod99)
   prodDens <- density(prod, n=365, from=1, to=365)

# Plot the density if the Plot parameter is T.
   if(Plot==T)
   {
      plot(prodDens, ylim=c(0,0.0043), type="l", xlab="Day", ylab="Density")
      lines(dens97, col=2)
      lines(dens99, col=8)
      legend(-10, 0.0044, lty=1, col=c(2,8,1), legend=c("1997","1999","Combined"))
#     title("Production")
   }

# Return the production density.
   prodDens
}
```

E.5 Sales Delay Density (*DelayDens.ssc*)

```
#               SALES DELAY DENSTIY

#------------
# DESCRIPTION
#------------
# PURPOSE
# =======
# This function creates boxplots and density plots of the delay in selling a vehicle, if the
# input Plot parameter is set to T.

# OUTPUT
# ======
# A vector of delay times from the claims97 and claims99 data frames.
```

```
#---------
# PROGRAM
#---------
"fDlayDns" <-
function (Plot=T)
{
# Filter out records representing second or subsequent repairs.
   dup97 <- duplicated(claims97$VIN)
   dup99 <- duplicated(claims99$VIN)

# Define dif as the difference between sale and prod. Look for outliers in delay. Remove cases
# where sale is NA or < prod.
   dif97 <- (claims97$sale[!dup97] - claims97$prod[!dup97])[!is.na(claims97$sale[!dup97]) &
       claims97$sale[!dup97] >= claims97$prod[!dup97]]
   dif99 <- (claims99$sale[!dup99] - claims99$prod[!dup99])[!is.na(claims99$sale[!dup99]) &
       claims99$sale[!dup99] >= claims99$prod[!dup99]]

   difBox97 <- boxplot(dif97,plot=F)
   difBox99 <- boxplot(dif99,plot=F)
   difOut97 <- difBox97$stats[1]
   difOut99 <- difBox99$stats[1]
# This is the largest dif that is not an outlier

# Sale will be in error if it is NA, < prod (this covers 0) or if sale is too long after prod.
   delay97 <- (claims97$sale[!dup97] - claims97$prod[!dup97])[!is.na(claims97$sale[!dup97]) &
       claims97$sale[!dup97] >= claims97$prod[!dup97] &
       claims97$sale[!dup97] <= claims97$prod[!dup97] + difOut97]
   delay99 <- (claims99$sale[!dup99] - claims99$prod[!dup99])[!is.na(claims99$sale[!dup99]) &
       claims99$sale[!dup99] >= claims99$prod[!dup99] &
       claims99$sale[!dup99] <= claims99$prod[!dup99] + difOut99]

# Combine these two into one delay vector.
   delay <-c(delay97, delay99)
   maxDelay <- max(delay)
   delayDens <- density( delay, n=maxDelay+1, from=0, to=maxDelay )
# Plot the graphs if graph parameter is T (default).
   if(Plot==T)
   {
       boxplot(delay97, delay99, names=c("Study Year","Later Year"), style.bxp="old")
       title("Sales Delay Boxplots")
       plot(delayDens, ylim=c(0,0.025), type="l",
          xlab="Sales Delay Time (days)", ylab="Density", Page="New")
       lines(density(delay97, n=maxDelay+1, from=0, to=maxDelay), col=2)
       lines(density(delay99, n=maxDelay+1, from=0, to=maxDelay), col=8)
       legend(275, 0.025, lty=1, col=c(2,8,1), legend=c("1997","1999","Combinded") )
#      title("Sales Delay Time Density Plots")
   }

# Return the delay density for the combined distribution:
   delayDens
}
```

E.6 Random Samples With Continuous Production (*SmpCont.ssc*)

```
#  GENERATE RANDOM SAMPLES OF PRODUCTION, DELAY AND FAILURE TIMES: CONTINUOUS PRODUCTION

#------------
# DESCRIPTION
#------------
# OVERVIEW
# ========
# This function simulates vehicle production, sales, and component failure.
# This function assumes that the observation period starts on 1 January of a year.

# INPUTS
# ======
# yrProdn = the number of vehicles produced in one year. The sample size will be this figure
```

```
#  multiplied by the length of the observation period.
# wrnty = the length of the warranty in years.
# obsPer = the observation period in years, simulating the time during which the data is
#  collected.
# beta = the Weibull distribution shape parameter used to generate a sample of failure times.
# eta = the Weibull distribution scale parameter.

# OUTPUT
# ======
# A matrix containing estimates of the two Weibull parameters obtained from the sample
   generated
# in this function.

#---------
# PROGRAM
#---------
"fSmpCont" <-
function(yrProdn, wrnty, obsPer, beta, eta)
{

# GENERATE PRODUCTION TIMES
# =========================
# Note: prodDens and cumProd are created in the script simulCont.ssc once and are available in
# this function. This saves about 1 hour of processing time for the 1000 simulations called in
# the script simulCont.ssc, which takes about 1 hour as is.

# Provided that obsPer is at least 1 year, randomly select one year's production days.
   if(obsPer >=1)
   {  prod <- round( approx(
         cumProd, prodDens$x, xout=runif(n=yrProdn, min=min(cumProd), max=max(cumProd)))$y )

# If obsPer > 1 year, randomly select production days for subsequent years. Need to add 365 to
# subsequent years.
      yrsLeft <- obsPer - 1
      while(yrsLeft >= 1)
      {
         nextYrProd <- round((obsPer-yrsLeft)*365.25) + round( approx(
            cumProd, prodDens$x, xout=runif(n=yrProdn, min=min(cumProd), max=max(cumProd)))$y
   )
         yrsLeft <- yrsLeft -1
         prod <- c(prod, nextYrProd)
      }
   }

# If observation period includes a decimal part, need to generate a part year's production.
   if(obsPer-floor(obsPer) > 0)
   {
      rem <- round( (obsPer-floor(obsPer))*365.25 )
      partYrProd <- round( approx(cumProd[1:rem], prodDens$x[1:rem],

      xout=runif(n=yrProdn*cumProd[rem],min=min(cumProd[1:rem]),max=max(cumProd[1:rem])))$y)
      prod <- c(prod, partYrProd)
   }

# Compute the sample size over the entire observation period.
   sampSize <- length(prod)

# CONVERT INPUT PARAMETERS FROM YEARS TO DAYS
# ===========================================
   wrnty <- round(365.25*wrnty)
   obsPer <- round(365.25*obsPer)

# GENERATE SALES DELAY TIMES
# ==========================
# Randomly select times from the delay densities.
   saleDelay <- round( approx(
      cumDelay, delayDens$x, xout=runif(n=sampSize, min=min(cumDelay), max=max(cumDelay)))$y
   )
```

```
# GENERATE FAILURE TIMES
# =====================
# Generate a random sample of failure times based on the Weibull distribution.
   samp <- round( rweibull(sampSize, shape=beta, scale=eta) )

# PUT THESE FIELDS INTO A DATAFRAME
# =================================
# Create a data frame with fields: ID, component's age at failure (compAge), failure/censored
# status (fail=1), vehicle's age at failure (vehAge) and component's age (age), be it censored
# or observed, at the end of the warranty, production day (prod), the time it takes to sell
   the
# vehicle (saleDelay), and calendar time (calTime). The field "age" will be set to wrnty, and
# adjusted to a part's age if it fails before the warratny ends.
   sampleDF <- data.frame(cbind(ID=1:sampSize, compAge=samp, status=0, vehAge=samp,
       age=min(wrnty,obsPer), prod, saleDelay, calTime=prod + saleDelay + samp))

# Identify and count records where failure occurs within the warranty and observation periods.
   inPer <- sampleDF$vehAge < wrnty & sampleDF$calTime <= obsPer
   numFailed<- count.rows(sampleDF[inPer,])

# For parts that fail within the warranty and observation period, set status to 1 and age to
# compAge, which is be the age at failure.
#  if(numFailed>0)
   {
       sampleDF$status[inPer] <- 1
       sampleDF$age[inPer] <- sampleDF$compAge[inPer]

# Set newRecords to sampleDF for use in the while loop below.
       newRecords <- sampleDF
   }

# Set count to 0, which is to be used in an escape clause.
       count <- 0

# REPLACE FAILED PART AND GENERATE RECORD FOR REPLACED PART
# ========================================================
# To be done while there are failed parts on vehicles still in warranty and observation
   period.
   while(numFailed > 0)
   {
# Generate new records.
       newSamp <- rweibull(numFailed, shape=beta, scale=eta)
       newRecords <- data.frame(cbind(ID=newRecords$ID[inPer],
           compAge=newSamp, status=0,
           vehAge=newRecords$vehAge[inPer]+newSamp,
           age=min(wrnty,obsPer) - newRecords$vehAge[inPer],
           prod=NA,
           saleDelay=NA,
           calTime=newRecords$calTime[inPer] + newSamp ) )

# See which of these new records are within warranty and observation periods, and count.
       inPer <- newRecords$vehAge < wrnty & newRecords$calTime <= obsPer
       numFailed <- count.rows(newRecords[inPer,])

# For parts that fail within the warranty and observation period, set status to 1 and age to
# compAge, which is be the age at failure.
       if(numFailed>0)
       {
           newRecords$status[inPer] <- 1
           newRecords$age[inPer] <- newRecords$compAge[inPer]
       }

# Add the new records to the existing ones.
       sampleDF <- rbind(sampleDF, newRecords)

# Escape clause if there is a problem.
       count <- count + 1
```

```
      if(count==50)
         stop(message="Error: Generated 50 new samples already.")
   }

# Fit a Weibull distribution to the sample.
   WblMdl <- survReg(Surv(age + 1, status, type = "right") ~ 1, data = sampleDF,
      na.action = na.exclude, dist = "weibull", scale=0,
      control=list(maxiter = 100, rel.tolerance=1e-004, failure=1))

# Return the Weibull parameters. The shape parameter (beta) is 1/WblMdl$scale and the scale
# parameter (eta) is exp(WblMdl$coeff).
   matrix(c(1/WblMdl$scale, exp(WblMdl$coeff) ), nrow=1, ncol=2)
}
```

E.7 Random Samples With 1 Year's Continuous Production (*SmpCont1yr.ssc*)

```
# GENERATE RANDOM SAMPLES OF PRODUCTION, DELAY AND FAILURE TIMES: 1-YEAR CONTINUOUS PRODUCTION

#------------
# DESCRIPTION
#------------
# OVERVIEW
# ========
# This function simulates vehicle production and sales, and assigns a failure time for a
# component.
# This function assumes that the observation period starts on 1 January and is >=1 yr.

# INPUTS
# ======
# yrProdn = the number of vehicles produced in one year. The sample size will be this figure
# multiplied by the length of the observation period.
# wrnty = the length of the warranty in years.
# obsPer = the observation period in years, simulating the time during which the data is
# collected.
# beta = the Weibull disrtibution shape parameter used to generate a sample of failure times.
# eta = the Weibull distribution scale parameter.

# OUTPUT
# ======
# A matrix containing estimates of the two Weibull parameters obtained from the sample
  generated
# in this function.

#---------
# PROGRAM
#---------
"fSmpCon1" <-
function(yrProdn, wrnty, obsPer, beta, eta)
{
# GENERATE PRODUCTION TIMES
# =======================
# Note: prodDens and cumProd are created in the script simulCont.ssc once and are available in
# this function. This saves about 1 hour of processing time for the 1000 simulations called in
# the script simulCont.ssc, which takes about 1 hour as is.

# Provided that obsPer is at least 1 year, randomly seclect one year's production days.

   prod <- round( approx(
      cumProd, prodDens$x, xout=runif(n=yrProdn, min=min(cumProd), max=max(cumProd)))$y )

# CONVERT INPUT PARAMETERS FROM YEARS TO DAYS
# =========================================
   wrnty <- round(365.25*wrnty)
   obsPer <- round(365.25*obsPer)

# GENERATE SALES DELAY TIMES
# ========================
# Randomly select times from the delay densities.
```

```
   saleDelay <- round( approx(
      cumDelay, delayDens$x, xout=runif(n=yrProdn, min=min(cumDelay), max=max(cumDelay)))$y )

# GENERATE FAILURE TIMES
# ======================
# Generate a random sample of failure times based on the Weibull distribution.
   samp <- round( rweibull(yrProdn, shape=beta, scale=eta) )

# PUT THESE FIELDS INTO A DATAFRAME
# =================================
# Create a data frame with fields: ID, prod, saleDelay, fail, status, calTime (calendar time).
# Note: As observation is for 2 years, all vehicles will be within the 3 year warranty, so
   there
# is no need to keep track of the vehicle's age.

   sampleDF <- data.frame(cbind(ID=1:yrProdn, compAge=samp, status=0, vehAge=samp,
      age=min(wrnty,obsPer), prod, saleDelay, calTime=prod + saleDelay + samp))

# Identify and count records where failure occurs within the warranty and observation periods.
   inPer <- sampleDF$vehAge < wrnty & sampleDF$calTime <= obsPer
   numFailed<- count.rows(sampleDF[inPer,])

# If there are failures during the warranty and observation period, set status to 1 and age to
# compAge, which will be the age at failure.
#  if(numFailed>0)
   {
      sampleDF$status[inPer] <- 1
      sampleDF$age[inPer] <- sampleDF$compAge[inPer]

# Set newRecords to sampleDF for use in the while loop below.
      newRecords <- sampleDF
   }

# Set count to 0, which is to be used in an escape clause.
      count <- 0

# REPLACE FAILED PART AND GENERATE RECORD FOR REPLACED PART
# ==========================================================
# To be done while there are failed parts on vehicles still in warranty and observation
   period.
   while(numFailed > 0)
   {
# Generate new records.
      newSamp <- rweibull(numFailed, shape=beta, scale=eta)
      newRecords <- data.frame(cbind(ID=newRecords$ID[inPer],
         compAge=newSamp, status=0,
         vehAge=newRecords$vehAge[inPer]+newSamp,
         age=min(wrnty,obsPer) - newRecords$vehAge[inPer],
         prod=NA,
         saleDelay=NA,
         calTime=newRecords$calTime[inPer] + newSamp ) )

# See which of these new records are within warranty and observation periods, and count.
      inPer <- newRecords$vehAge < wrnty & newRecords$calTime <= obsPer
      numFailed <- count.rows(newRecords[inPer,])

# If there are failures during the warranty and observation period, set status to 1 and age to
# compAge, which will be the age at failure.
      if(numFailed>0)
      {
         newRecords$status[inPer] <- 1
         newRecords$age[inPer] <- newRecords$compAge[inPer]
      }

# Add the new records to the existing ones.
      sampleDF <- rbind(sampleDF, newRecords)
```

```
# Escape clause if there is a problem.
      count <- count + 1
      if(count==50)
          stop(message="Error: Generated 50 new samples already.")
    }

# Fit a Weibull distribution to the sample.
   WblMdl <- survReg(Surv(age + 1, status, type = "right") ~ 1, data = sampleDF,
      na.action = na.exclude, dist = "weibull", scale=0,
      control=list(maxiter = 100, rel.tolerance=1e-004, failure=1))

# Return the Weibull parameters. The shape parameter (beta) is 1/WblMdl$scale and the scale
# parameter (eta) is exp(WblMdl$coeff).
   matrix(c(1/WblMdl$scale, exp(WblMdl$coeff) ), nrow=1, ncol=2)
}
```